Statistics Canada
Agriculture Division

WORKING PAPER #34

Utilization of Document Imaging Technology By the 1996 Canadian Census of Agriculture

Mel Jones and Ivan Green
Statistics Canada

February 1998

Catalogue no. 21-601-MIE98034

The responsibility for the analysis and interpretation of the data is that of the authors and not of Statistics Canada.

© Minister of Industry, Statistics Canada, 1998. All rights reserved. No part of this paper may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise.

Utilization of Document Imaging Technology

By the 1996 Canadian Census of Agriculture

Mel Jones and Ivan Green
Statistics Canada
February 3, 1998

Abstract

Processing of the Canadian Census of Agriculture relies heavily upon access to its questionnaires. The documents are required within many stages of collection and processing to ensure delivery of quality products to agricultural clients. Historically, access to the physical documents has proven problematic in terms of timely delivery, concurrent access, management, tracking, storage and cost. As early as 1984, Census management recognized the potential of document imaging technology to improve data quality and timeliness. However, in the 80s it took little effort to recognize that the technology of the day was not sufficiently mature or affordable for the Census. In 1996, document imaging became a successful reality for the Canadian Census of Agriculture. This paper describes the experiences, the issues and the expectations of the many different players involved in the implementation.

Introduction

Implementing document imaging technology for the Canadian Census of Agriculture involved significant expense and risk taking. In the early stages of the project, healthy discussions took place over conventional methods and competing technologies. The first section explains the issues embedded in the discussions and sheds light on the direction chosen. It also contrasts the drawbacks of the historical manual methods with the expected benefits of document imaging.

The technical challenges associated with implementation of document imaging were many. The second section touches briefly on several of the challenges and explains how key technical issues of document imaging were identified and satisfied. It closes with observations on some obvious and not-so-obvious reasons why the introduction of the technology succeeded.

Section three contains a management perspective on the decision to use imaging technology and offers some possibilities for its continued use and evolution in Statistics Canada.

Section I: Why Imaging?

In accordance with agency policy, release dates for publications are made available in advance to the public. When declared, the release dates do not change. Consequently, statistical production adheres to very strict deadlines to meet release dates.

Time, and fiscal constraints, in addition to the need to maintain or improve quality have fueled a drive to become more efficient. As a result, increased emphasis is being placed on automation and innovation. Expensive manual processes are prime targets.

Prime targets for the Canadian Census of Agriculture were (a) data capture and (b) the delivery of physical questionnaires throughout the production workflow. Feasibility studies to automate were conducted for both.

For data capture, an investigation of intelligent character recognition (ICR) technology was carried out. At the time, the maturity of ICR technology for Census purposes was in question and there was a lack of time to test respondent reaction. To use it would have significantly increased respondent burden. This is because a change from unconstrained hand-written numeric responses to ICR-friendly, constrained numerics would have been required. Since greater than 90% of the information sought by the Census is numeric, a complete redesign of the questionnaire would likely have been necessary. After the feasibility study the efficacy and cost-effectiveness of recognition software was also in question.

The research into Imaging and ICR technology lead to a very interesting and challenging situation. In Canadian political terms you would say the project

management team fell into 3 camps; the conservative, the progressive-conservative and the progressive. The conservatives felt that implementing imaging by itself was too high risk because:

- We were 2 years away from the census. For a quintenial project, it was too high risk to test and implement imaging in that time frame.
- The cost estimates varied widely from affordable to way out of our league.
- The department had no experience with imaging and ICR technology. It would be safer to prove the technology on smaller and more frequent surveys.
- The staff would be totally dependent on machines. If the system went down, staff were out of work.
- The project had many other challenges including:
    - The addition of a new Progress of Seeding Survey. This was a CATI follow-up survey to over 40 % of census respondents.
    - The complete re-engineering of processes and migration from the main frame to mid-range technology.
    - a new system and methodology for the register of farms.

The conservatives understood the value and benefits of imaging but felt the risks were too high for the 1996 Census of Agriculture. The traditional manual method (no imaging route) was safest and proven.

The progressive camp believed the greatest savings could be realized by implementing imaging and ICR together. They believed imaging without ICR was not cost effective. They were not only convinced that imaging could be done but that ICR technology was then sufficiently mature for census purposes.

The project manager's original proposal was to implement imaging without ICR. The progressive-conservative group supported it. Risks associated with ICR were unacceptable given the timeframe faced by the project. This camp however, felt imaging was feasible with a manageable amount of risk. While they had heard of many disastrous imaging projects, they had also seen enough working applications in other locations to convince them that the risks were worth the benefits.

There was no right or wrong in the debate. Each group had valid reasons for their stand. Once the decision was made to implement imaging without ICR, the decision was fully supported by the entire project team.

Delivering electronic images of questionnaires to the desktops of editors and validators was now the goal.

If a manual questionnaire delivery system had been implemented for 1996, it would have suffered from the same shortcomings as its predecessors:

- ***Timeliness.*** Delivery of questionnaires to production staff would take hours and in some instances, days.
- ***Contention.*** This is another timeliness issue. Often, validators and editors would require concurrent access to a questionnaire from a large farming operation. Large farming operations are the top contributors to estimates for many Census variables. Delivery would be conducted in a serial manner because copies could not be made. Having duplicate copies was considered a security risk and was not permitted. To keep track of documents the questionnaires were filed and pulled again between each operation or request.
- ***Tracking.*** Maintaining an awareness of the location of each document would require development of an automated tracking system. A manual record of receipt via bar-code readers would be needed at each stage of production. This type of system would impose a serial movement of questionnaires (and data) through the processes.
- ***Complex storage.*** An elaborate system of folders, boxes and shelving would be required to support high-volume retrieval and storage of documents.
- ***High costs.*** Staffing, shelving, storage space, security and development of a questionnaire distribution and tracking system all contributed to an expensive, comparatively slow manual system.
- ***Frustration.*** Validators had to make lists on paper of needed questionnaires and then submit these to the library. In a matter of hours they normally would get their request back. However, many would be missing due to contention, filing errors, etc. Several requests would have to be made before all documents were located.
- ***Filing errors.*** Filing and refiling resulted in errors. Periodically the library operation would be shut down to do a systematic and complete check of the library.

Intuitively, a questionnaire imaging system addressed the shortcomings of its manual counterpart:

- ***Timeliness.*** Delivery of questionnaire images to production staff would take seconds rather than hours or days.
- ***Contention.*** All users could examine the same document concurrently. This would create opportunities for ***parallel*** approaches to processing.
- ***Tracking.*** No manual questionnaire tracking system was necessary throughout processing workflows. In addition people were spared the effort of documenting receipt and release of questionnaires.
- ***Complex storage.*** A temporary storage would still be required. However, the system of storage would be comparatively simpler.
- ***High costs.*** Since imaging had never been conducted on a large scale before at Statistics Canada, there was no benchmark with which to compare the Census of Agriculture. As a consequence, it was not clear as to whether imaging was more or less expensive than a manual questionnaire delivery system.

Management believed that the potential benefits warranted a closer look.

In addition to those benefits listed above, management expected the following:

- no need to micro-film the documents after processing
- reduction in manual operations, particularly the library staff
- immediate access to documents would make staff more efficient as they could find a problem and solve it immediately. No need to make a list and then find something else to do until the documents arrive
- reduction of paper handling by all clerical and validation staff would be realized. Delivery of electronic images would eliminate much of the need to handle paper.
- remove frustration caused by delays in receiving or waiting for questionnaires and missing questionnaires
- job enrichment: staff would find previous mundane jobs satisfying and challenging
- immediate access to problem questionnaires meant staff could rely more on information than intuition. In order to avoid waiting for the paper document some staff would guess at the problem and the solution.
- immediate access would allow staff to look at more problems and thereby improve the quality of the estimates, most noticeably at the lower geographic levels.

Section II: Imaging Issues

When the investigation of document imaging began, answers to the following questions were sought:

- What is document imaging?
- What components of imaging are necessary to meet our requirements?
- Will it work in our technical environment?
- What is a rough estimate of the costs?
- What issues must be addressed for successful implementation?

As a first step towards answering these questions, Census representatives began discussions with the private sector. Industry was provided statements of requirements, volumes and samples of the 1991 questionnaire which was considered to be representative of the (as yet to be developed) 1996 version. In return, industry representatives provided cost estimates and descriptions of potential solutions.

Estimates ranged from $500,000 Cdn. to almost $3,000,000. The lower estimates were affordable, the upper weren't.

The solutions delivered with the estimates were somewhat confusing. They addressed some problems that weren't clearly applicable to Census and omitted others that possibly were. However, despite some minor confusion, Census representatives began building a clearer picture of what the issues were.

Some emerging (and very expensive) issues, which concerned Census, were:

- Indexing,
- Quality control,
- Throughput of scanners using legal-sized, 70m Vista paper; the Census questionnaire,
- Required adjustments to the layout of the questionnaires to make them ***image-friendly***, and,
- The quality of the images.

To build a sufficiently complete list of issues, seminars on document imaging at Association of Information and Image Management (AIIM) trade shows were attended and a visit with imaging researchers at the United States Bureau of the Census (USBC) was made.

USBC representatives brought to light the critical issue of ***document integrity***. USBC questionnaires are multiple-page booklets. Scanners deal with one sheet of paper at a time so imaging a multi-page document may require cutting to allow access the individual pages. Of primary concern to USBC researchers was the potential for pages of a multi-page document to be lost or become out of sequence during movement to or from the scanner; i.e. loss of document integrity. At the time, research of page-turners (rather than cutters) was being undertaken by USBC. Since the Census of Agriculture's questionnaire is multi-page it was believed that this was an issue for Statistics Canada as well. Given the relatively small volume (280,000) of questionnaires for the Census, page-turners did not appear to be a cost-effective solution.

One of the first solutions to come to mind was to print a unique questionnaire identifier on each page of each questionnaire. Discussions were held with internal experts and with large print shops. The solution was indeed feasible, however, not cost-effective. Estimates provided were close to $400,000 Cdn., adding approximately $1.43 to the cost of printing each questionnaire and more than tripling the budget for printing. The chance of error during sequencing process still existed. A second solution considered was to have the scanner emboss a serial number on the bottom of each page that it scanned. This was a suitable solution for questionnaires dropped ***after*** they had been scanned; it still didn't address those dropped on their way to the scanner. More work would have to be done.

Investigation of document imaging continued; trade shows illuminated the following additional issues:

- Size and quality of monitors,
- Video cards and functionality,
- Network capacity,
- Storage capacity,
- Scanner functionality,
- Technical support, and
- Integration with in-house applications.

At this point a reasonably complete list had been compiled, but it still wasn't clear how it all applied to the Census of Agriculture. There was a concern that the Census application might have some abnormal imaging issues to address. For example, most scanners are rated on single-sheet, letter-sized, 30-weight paper. Census uses 16 page booklets of legal-sized 70M Vista paper. Management agreed to fund a small pilot ($20,000 Cdn) to clarify the Census issues.

A local firm was contracted to set up a basic imaging system on-site. They worked closely with internal systems and methodology experts. The objectives were:

- To investigate workflow methods and procedures to fully operationalize imaging in terms of questionnaire receipt, preparation, scanning, reassembly, storage, retrieval and quality control.
- To become familiar with industry standard image hardware and software capabilities, and
- To become familiar with performance of the technology with respect to Census documents.

The pilot system was composed of:

- Two different scanners (Fujitsu and Bell + Howell),
- A Hewlett-Packard server running UNIX and LAN-Manager 2.1,
- Four PCs with 2 MB video cards,
- 17" and 21" color and Grey-scale monitors,
- TCP/IP
- Ethernet 10 Base T, UTP subnet
- Watermark TIF image viewers, and
- A sample of 2,000 x 16 legal-size page questionnaires from the 1991 Census.

The pilot lasted three days and yielded the following results.

- Increased political acceptance of the technology. Observation of questionnaires being scanned, the quality of images and the functionality of image viewers boosted confidence and reduced skepticism.
- Due to Statistics Canada's adherence to interface standards, there were no unexpected surprises arising from the technical infrastructure.
- The observed throughput of the scanners was considerably less than rated, but was still impressive.
- In the Census experience, equivalent models of properly serviced scanners from competing manufacturers are not equivalent regarding throughput. One of the scanners performed reliably, having relatively few double-feeds, skews and jams. The other scanner was very poor, literally destroying questionnaires; its use in the pilot was quickly discontinued.
- The functionality of scanners was quite extensive in terms of reading bar codes, indexing, counting, etc.
- The number of scanners of this type and the staffing required for Census volumes was determined,

- The quality of the images was acceptable (200 dpi),
- The average size of the images (produced by the pilot scanners) was determined; this information would be useful in simulating network traffic and in determining image storage requirements,
- The weight and color of the paper used for the questionnaires (Vista 70m opaque) was acceptable, as no markings from one side of a sheet of paper bled through to the other.
- The layout of the 1991 questionnaire would be unacceptable for imaging. However, adjustments would be minor and would have no impact on the respondent or on printing costs for 1996. No testing of respondent reaction to the image-friendly questionnaire was necessary.
- The functionality, look and feel of a Watermark viewer would meet Census requirements, this would allow a crude estimate of viewer costs and serve as a reference point for specification.
- Document integrity was not an issue. Concern existed over cut documents being dropped and pages lost or misplaced; loss of integrity. Tests demonstrated that batches of cut questionnaires tended to stay together when dropped. This was due to static electricity and folds in the pages. Nevertheless, the risk of lost or misplaced pages still remained. Measures were taken to further reduce the risk. They included keeping the batches small (30 or less documents), storing batches in envelops, minimizing time out of the envelops, reducing the distance between cutter and scanner, using hand-writing style and ink color to reassemble and as a last measure of insurance, using either the images or the data in the structured database to reassemble dropped documents.
- The less expensive 17" color monitors and video cards were acceptable (at 1024x768) for at least half of the Census production activities. The quality of the image was often ***better*** than the original; due to the fact that images can be enlarged. The grey-scale monitors provided a clearer display, but not significantly. The color monitors were considered easier to deploy after production to other activities.

The results of the pilot test paved the way towards implementation. What remained was:

- Quality control strategy
- Indexing strategy
- Determine network requirements
- Determine integration requirements
- Determine scanner requirements
- Determine viewer requirements
- Selection and acquisition of monitors and video cards
- Acquire storage technology
- Develop the workflow and control systems for the scanning operation
- Develop an RFP
- Implementation

Throughout all of the remaining activities the contributions made by representatives from the Agency's Informatics User Services Division, the Mid-Range Support Group, Business Survey Methods Division and the System Development Division proved to be of critical importance.

Bandwidth: A major concern regarding image production loads was the network. A few years before the project, the Agency had rewired it premises to 10 BaseT UTP. Despite this the wiring had proven to be vulnerable to movements of large files. Assurance that it would support the concurrent activities of 100 end users and a high volume scanning operation was needed. The fact that no benchmark existed within the Agency for imaging volumes posed a problem. Fortunately, key task managers from the previous Census were still available and provided information about the volume of questionnaires drawn on a daily basis using the manual system. It was anticipated that imaging volumes would exceed that of the hardcopy documents. For testing purposes, the volume was tripled to represent user traffic. It was added to the expected volume of scanning traffic (total volume of questionnaires divided by the total number of days available to scan) to represent the total daily imaging load on the network. The total load was divided by 6.5 hours (average level of activity in 7.5 hour day) to yield the average load at any point in time during production. Using this formula and the information regarding image sizes obtained from the pilot, a series of network tests were conducted.

Experts from the Agency's Communications Services Center (CSC) were called to assist in the tests. After consultation, CSC provided advice on how to effectively conduct tests to yield the most meaningful results. Simulating the megabits per second demand, i.e. a workstation pulling and pushing images to and from a server over the network was insufficient to represent 100 users. Simulation of the ***distribution*** of the demand was also important. If only one workstation pulled and another pushed, real-life network collisions were not likely to occur. Many workstations had to be employed if the tests were to be meaningful.

Prior to conducting the tests, CSC configured an isolated subnet. This was done to eliminate noise so that the loads being measured were attributable to movement of images. Twenty workstations were used. On each was a program that either copied a file to or from a server on a specified time interval. CSC used network analyzers called "sniffers" to observe the traffic.

Based on CSC recommendations, CSC acquired an Ethernet Routed Switch with a 100 mbs capacity to support production loads. Representatives from the Technical Support Group installed the switch, isolating the Census users from the rest of the Agency. The purpose was two-fold: (1) to protect the imaging subnet from performance degradation due to unanticipated fluctuations in external LAN traffic and (2) to protect external LANs from the Census imaging loads.

Census users were organized into three workgroups on the Ethernet switch with a 10 mbs connection. The server was linked to the switch by a 100 mbs fiber-optic cable. Census funded acquisition of a 100 mbs Fiber Direct Data Interchange (FDDI) card to interface the server with the network. The Mid-Range Support Group investigated solutions, made a recommendation and, after consultation, acquired and installed the FDDI card.

After the *image subnet* was implemented, CSC returned to conduct further tests to ensure that expectations had been met. They had been.

Monitors: Since the imaging system exists to serve the end-user and most often the only awareness of the imaging system is based on what is on the screen, the monitor is arguably the most critical component of an imaging system. It is certainly one of the most expensive. During the pilot test no significant difference in the appearance of Census images was observed between the 17" color and the 19" Grey-scale monitors. Since the monitors were to be used for more than imaging applications, despite the higher cost, it was decided that color monitors would be acquired. What remained was to determine (a) the size or sizes of monitors that meet the Census needs and (b) the video card requirements.

Some stages of the Census production workflow require concurrent observation of two or more legal-sized pages. It wasn't known if a 17" monitor with an appropriately configured video card would or would not be sufficient. The cost difference between a 17" and a 21" is very significant. It had to be determined if the expensive larger monitors were actually required.

At the time there were several 17" monitors being used by Census staff that could be used for an imaging evaluation. However, larger monitors were not available. Representatives from the private sector were contacted for some 21" *loaners*.

Technically oriented members of the Census production staff volunteered to evaluate monitors and video cards. Large and mid-sized monitors were rotated amongst the group for a period of three weeks. It was determined that in spite of *virtual desktop* capability and a resolution of 1280x1024, the 17" monitors were insufficient to meet all the Census requirements. Virtual desktop is a feature of video cards. It allows the user to specify desktops that are larger than the physical dimensions of a monitor's display surface. The areas of the virtual desktop that are not visible can be brought into view by movement of the mouse. The virtual desktop was considered essential for some of the activities to be performed on the 21" monitors. Based on the research, the quantity of medium and large monitors was determined; the budget was next.

The pricing, physical dimensions and performance of monitors vary considerably. For example, it was necessary to ensure that the 21" monitors would comfortably fit on the furniture; some didn't. Perhaps more importantly, the quality of the display in terms of sharpness, contrast, brightness and consistency of image on the entire display surface varied considerably among comparatively configured and priced monitors.

The Agency maintains standing offers (SO) with vendors to facilitate acquisition of informatics technology. On the SO list were offerings from several vendors. It was in the project's best interest to purchase from the SO if any of the monitors on the list met the requirements. A representative from the Agency's Workstation Support Group provided loaners of all the 21" monitor types on the list. The monitors were simultaneously driven by the same workstation and displayed the same image. Prices ranged from $900 Cdn. to $3,000. From a distance, the least expensive monitor looked good. However, closer inspection revealed blurring and fading of colors at the edges of the display. Some flicker at a resolution of 1024x1280 was also noticed. The higher priced monitors performed very well but were more expensive and lower in quality than one of the models previously evaluated by Census staff. Arrangements were made with the vendor to have

their product added to the SO. (The desired video cards were already there.) These monitors were included in a purchase of 120 Pentiums destined for production.

During production it was noticed that many of the end-users didn't take advantage of the virtual desktop feature of the video cards simply because they weren't comfortable with it. In addition, many users didn't use the higher resolutions supported by the monitors and video cards. Many reverted to a resolution roughly equivalent to that on their former 14» SVGA monitors. Perhaps the clear, small text supported by the technology is too small to read, especially for people with visual impairments.

Quality Control: Many imaging systems have workstations and staff responsible for the quality of the images made available to production workflow. QC measures can significantly affect the cost of document imaging, often requiring expensive equipment, consuming salary and potentially introducing a bottleneck.

The bottleneck was of particular concern. Most workflows involving document images are image driven. The Census workflow was driven by data. This means that when an editor or analyst had access to the captured data on the structured database, they were ready to begin work. If the corresponding questionnaire image had not yet been delivered, they would either continue without its benefit (defeating the purpose) or wait until it was available. For the Census, it was imperative that images were delivered either ***before or at the same time*** that structured data was. To inspect (in a timely manner) the daily volumes of images generated by the scanning operation without causing delays in production workflow would have required an expensive QC operation.

Discussions were held with Census management, key staff, methodologists and systems personnel to look for ways to circumvent the problem. As long as the page was legible, it wasn't of concern to the user if the rest of the document was corrupt. Most often, end-users only looked at a portion of one page of each 16-page document they requested. A methodologist, reviewing information (statistics on double-feeds, skews, tears and folds) obtained from scanner evaluations, determined the expected number of times an end-user would encounter an unacceptable image and request a rescan. They were negligible. This was due to the consistently high quality of the images generated by the scanner. As a consequence of this investigation, it was decided that QC operations would be performed after, rather than during the time critical production period. This decision was made easier due to the fact that to request a rescan, required the end-user to simply click on a ***rescan button***.

As a result of some creative thinking and closely focusing on the need rather than implementing conventional technical solutions, the amount and cost of QC was significantly reduced. No additional workstations were required and staff was tasked with QC during their ***slack time***.

Indexing: As with other types of database entities, one or more keys are associated with an image to support timely retrieval by the end-user. Some applications employ optical character recognition (OCR) to automatically extract identifying information from an image. The information would then be used to build indexes.

Since the identifying questionnaire information was in unconstrained handwritten form (rather than typewritten or printer generated), OCR was not a feasible solution for the Canadian Census of Agriculture. ICR had already been ruled out. An alternative was to build an interactive indexing application. The implementation would require staff to view selected pages from each questionnaire image and data capture the identifying information. This data would then be used to build indexes to the imaged documents. Two drawbacks existed: (1) index stations would increase the cost of imaging and (2) another potential bottleneck would be introduced prior to the production workflow.

Fortunately, the necessary indexes could be obtained elsewhere and at no additional cost. All appropriate identifiers referenced the database, which housed the data-captured questionnaire information. A link was required between the images and the database. Representatives were called in from the Agency's System Development Division to seek a solution.

It was known that scanners could read barcode labels and automatically register the codes as indexes to images. Based on data capture plans, the barcodes were to be affixed to the questionnaires at the time of keying and were to be captured in addition to the respondent data. The codes would be used as unique identifiers on the database.

Given this scenario, systems people proposed that a program be written to: (a) access new images on the image base to obtain a barcode number and the image base internal identifier and (b) using the barcode, find

the matching record on the database and store the image base internal identifier within it. This would permit end-users to access a database record using any of the predefined keys and then select the corresponding image from the image base using the internal image identifier. This was an effective, yet relatively uncomplicated solution.

The Shape of the System: What was once very cloudy regarding the imaging solution was becoming clear. Since QC applications, indexing applications and embossing of questionnaires were not required, the essential software components that appeared to remain were:

- Scanner management software
- Image base management
- Image viewers
- Bridges between the image system and the Census application (integration).

The hardware components were:

- Scanner(s)
- Server
- Storage device(s)
- Network infrastructure
- Workstations, monitors, video cards

The paper component (the questionnaire) had been dealt with in terms of color and had been made *image-friendly*.

Scanner Management Software: Much had been learned about scanner functionality and performance through the pilot test, trades shows and through discussions with industry. However, the evaluations using Census paper had not been applied to an industry leader: KODAK. This appeared to be a significant omission because KODAK was just as probable of being in the winning RFP as the other scanner manufacturers. Additionally, (other than using a barcode as a key) the critical details of how a scanner was to manage Census documents had not been addressed.

Given the higher-end capacity of some of KODAK's models relative to the scanners in the pilot, fewer staff might be required in the scanning operation. A meeting between a KODAK representative and a Census manager tasked with operationalizing the scanning operation was held. An on-site loan of an ImageLink 500 was arranged. Census representatives, a methodologist, micro-computer support staff, systems people and KODAK representatives worked closely for a month. In addition to gaining a familiarization of the functionality and performance characteristics of a KODAK, comparisons could be made to offerings by other manufacturers used in the pilot test. A sense of which scanner technology would best suit Census needs was occurring.

Tests were designed to address as many aspects of scanning for Census as could be imagined. The operations manager, her assistant and a methodologist:

- focussed on the throughput of the scanner,
- the quality of images,
- determined whether a white light was sufficient to scan both colors of the questionnaire and
- learned what was involved in operating and maintaining the scanner.

Critical outputs from this exercise were:

(1) scanner functional specifications to be included in the RFP,
(2) benchmark figures (occurrence of double-feeds, skewed images, jams),
(3) average sizes of acceptable KODAK images (these were consistent in size with images produced in the pilot test),
(4) a stronger sense of the staffing issues,
(5) a discovery that the KODAK transport mechanism worked better with half rather than full batches (15 versus 30) of Census documents, and
(6) a discovery that barcode labels must be very precise or the scanner will not recognize them. All bar-code labels are not of equal quality. If the *hands-on* evaluation had not taken place, this critical discovery might have been made at a potentially crippling time.

Following this evaluation, scanning management began addressing the considerable challenges associated with operationalizing the scanning activities. This involved designing a workflow, which dealt with receipt of documents, sorting, scanning and eventual storage. An automated system to track the movement of documents throughout the normal workflow and through rescan requests was designed and specified. An Agency systems representative built it (Visual Basic), participated in the test activities and installed it. The scanning operation functioned very efficiently and very reliably in production.

During the KODAK evaluation, a senior Census Task Manager began managing the development and delivery of the RFP. He recruited the operations manager, her assistant, the Census technical liaison, a representative from Mid-Range Computers and a

representative from the System Development Division. The contents of the RFP were defined to be:

1. Background information about the Census,
2. Information on the technical infrastructure (server technology, OS, network protocol, desktop configuration, desktop OS…),
3. Scanner functionality and performance requirements,
4. Training requirements,
5. Integration requirements,
6. Technical constraints such as image sizes,
7. Image viewer functional requirements,
8. Support requirements,
9. Acceptance Criteria, and
10. Time constraints.

Issuing an RFP in the Canadian Federal Public Service is subject to many rules and regulations. Although necessary, the result is that at typical RFP can consume up to six months from delivery to Public Works to eventual declaration of a winner. Time was becoming a critical factor for the project, as production was less than eight months away. It is believed that the RFP process could have placed the project in jeopardy had the manager not kept in constant contact with Public Works, pressuring them to keep the Census RFP as a high priority.

Image Viewers: Options existed in the market; viewers could be acquired or custom ones built. Some industry representatives advised that considerable savings could be achieved if a custom one were to be built. Others advised against it because of proprietary issues. It was not known whether the winner of the RFP would build or utilize an existing viewer. Either way, the functional requirements had to be determined and included in the RFP.

The determination of requirements turned out to relatively simple. A Watermark image viewer was obtained and made available to selected Census staff. After a demonstration and some time to evaluate the software a walkthrough was held. During the exchange, the individuals simply identified what they felt was required, what should be modified and then prioritized it. They also indicated what they felt was not required. The results were then documented and included in the RFP. This ensured that the viewer requirements would be met whether off-the-shelf or custom-built technology was delivered by the contractor.

Storage Requirements: This was relatively easy to do. After examining samples of images obtained in the pilot and in the evaluation of the KODAK, an average image size of approximately 600K was determined. The size was multiplied by the number of questionnaires (along with a 25% insurance factor) to derive the total storage requirement.

Some risk was involved here, however. This was due to the fact that the RFP for storage technology had to be let before a winner of the imaging RFP was declared. The size of the image had a direct impact on three critical items:

1. ***Image quality***. The larger the image file, the better the quality. Care had to be taken that the images were acceptable.
2. ***Storage requirements***: When dealing with large image bases (4.8 million images) subtle changes in image sizes can significantly affect storage requirements. Adequate storage had to be available prior to production.
3. ***Performance***: The larger an image the greater the load on the network, the server and the workstation. All the preparatory performance tests had been done with an anticipated image size. A significant increase in image size may have had adverse affects.

Performance: A critical concern for the Census production managers was the speed at which a requested image is delivered to the desktop. It was potentially the difference between success and failure in Census imaging.

Minimizing delivery times was a challenge. This was due to the many layers of technology, which delivered images and competition for computing and network capacity by other processes.

Performance issues included:

- The desktop
  - Clock speed
  - RAM
  - Cache
  - Swap space
  - OS (16-32bit)
- The monitor and video card
- The Network
  - Capacity (10-100 mbs)
  - Protocol (TCP/IP, packet sizes…)
  - Topology (isolated, not isolated)
- Images
  - Size of images
  - Size of image base
  - Database technology
- Server
  - RAM

    - OS
    - Dedicated or supporting other application
- Storage devices
    - RAID caching
    - Jukeboxes
        - Optical medium density
- Imaging application
    - Viewers
    - Image Base Management
    - Indexes……

As is evident, holding the contractor to specified performance levels in production was not possible. This was due to the many environmental factors outside his control. However, it was still essential to determine that the performance of contractor's solution would not jeopardize the project.

A performance test was designed, described in the acceptance criteria of the RFP and implemented. In an isolated subnet, using the production server and storage technology, the contractor was required to do the following:

- Scan versions of the 1996 Census questionnaire.
- Create multiple copies of the images.
- Populate the image base to a level representing roughly half the eventual size it would attain in production.
- Using two workstations, retrieve volumes of images representative of anticipated production volumes in a specified time frame. This was staged, starting with one questionnaire, then 10, then 100…

Under this scenario, the contractor's solution was required to deliver the first page of a questionnaire within 5 seconds and subsequent pages of the same questionnaire in less than 1 second.

Of equal importance was the integration of the contractor's offering with internally developed production systems. In order to meet requirements, the contractor had to work with a systems development expert to clearly demonstrate that an image viewer (complete with an image) could be launched successfully from the Census production application.

After the tests, Census representatives were reasonably confident if a performance problem occurred that the vendor's solution would not likely be the cause.

On the Census side of the technology, steps were taken to optimize performance.

1. The network was configured, sized and properly tested.
2. A large RAID device was acquired to cache frequently requested images
3. Pentiums with sufficient RAM and disk were acquired
4. High performance monitors and video cards were acquired.

The net result was satisfactory delivery of images in production. A high level of availability was also experienced as downtime was minimal.

Why It Worked: When the move to document imaging began for Census of Agriculture, stories of imaging disasters abounded, no significant imaging applications existed within Statistics Canada and the people tasked with delivery had no experience whatsoever with the technology. A very expensive and visible failure could have happened for the Canadian Census of Agriculture.

A reflection provides some reasons why it succeeded:

- ***Statistics Canada's Informatics Organization***. Technical expertise for mid-range computers, networks, workstations and system development were available on a moment's notice to contribute to the project. This was just as critical during production as it was in the activities leading up to it.
- ***Statistics Canada's Technical Infrastructure***: For some time, the Agency has strongly emphasized adherence to standards. This set the stage for successful introduction of standards compliant technology into the environment.
- ***Methodologist***: A methodologist assisted team members in analyzing the technology and in producing meaningful statistics.
- ***Technical Liaison***: Today's informatics projects often require several different technical disciplines (mid-range computer specialists, network specialists, workstation specialists, applications developers, database specialists, product specialists…). The imaging project employed a technical liaison that coordinated activities and represented the Census with industry and technical groups.
- ***Turnkey System***: Rather than split the contract into say, Company #1 to generate images, Company #2 to manage images and Company #3 to provide image viewing services, a complete imaging system was purchased. This ensured that all pieces of the system would be interoperable. (i.e. images produced by the scanner would be compatible with the viewer). It also prevented

communication problems in the event that service was required. For the Census, there was only one contractor to call.

- ***Simplicity***: The absence of indexing stations and quality control stations reduced the complexity of the application. In addition, a minimal interface between the Census production application and the FileNet Imaging System reduced integration issues and correspondingly reduced integration problems.
- ***Hands-On Experience***: Sole reliance on literature and demonstrations would have been dangerous, particularly with scanner functionality. In addition, hands-on evaluation of monitors ensured that adequate quality was acquired; it would have been easy to spend too much or, disastrously, not enough. Wherever possible, Census staff acquired loaner technology and gained hands on experience using the Census documents and environment.
- ***Focusing on the Need***: Often, throughout the course of the project, it would have been easier to go with conventional solutions such as indexing stations and QC stations. This would have unnecessarily increased costs, complexity and risks.
- ***Perseverance***: Several times, throughout the course of the imaging project, it could have been terminated.
- ***Management Support and Personal Involvement***: It's widely known that many projects fail due to a lack of commitment by management. The imaging project didn't suffer that consequence; far from it. The assistant director and particularly the project manager publicly demonstrated support, fought the political battles and played a major role in shaping the eventual system.
- ***Relationship with the Contractor***: This was excellent. Keys to success were availability, accessibility and willingness to compromise. It's important to note that the project represented a foothold in the Canadian Federal Government for the contractor. Success was as important to the contractor as it was to the Census. This undoubtedly had an impact on the relationship.
- ***Many Contributors***: Many people were involved in shaping the imaging system. Project staff were involved in:
  - Assessing monitors, video cards, image viewers and scanners.
  - Sizing the network and building programs to test it
  - Defining the overall structure,
  - Building RFPs for production and pilot tests
  - Defining how images would be used in production
  - Others fought the unavoidable political battles.

  This broad level of involvement cultivated a greater sense of ownership and undoubtedly contributed to the success.
- ***Budget***: Acquisition, implementation and maintenance costs are not trivial. A healthy slush fund allowed for unforeseen adjustments to the implementation.
- ***Paper Questionnaires Weren't an Option***: Often, when technologies are introduced, the technology they are designed to replace still remains. Even if a compelling reason to change exists, unless people see it, they often continue going about their work using the tools they are most comfortable with. For the Census of Agriculture, the 1996 application represented a complete redevelopment. The manual questionnaire delivery system employed in 1991 no longer existed. Only in exceptional circumstances was a physical document drawn in 1996.
- ***Acceptance***: Production staff were highly satisfied with the quality of the images and functionality of the viewer. Annotation, zooming and scrolling were considered to be extremely productive features.
- ***Luck***: The project was fortunate to have many people on the team and in the service areas blessed with talent. Two or three key project people were remarkable.

Section III: Reflections and the Future

The goal of timely delivery images to the desktop was met. Acceptance and appreciation of the technology by all users and observers was positive to say the least. Imaging combined with an excellent set of edit and validation tools proved to be a powerful combination for users. The only question staff had was how did we process the census in the past without the images?

The expected benefits were realized and more. Use of images occurred in many processes and applications where it was neither planned nor anticipated. Some of the unexpected benefits were:

- Productivity improvements far exceeded the expected 10% level. In many clerical operations the image viewer automatically zoomed in on the questionnaire's problem module and the

correction was made on line. This proved to be several times faster than leafing through a paper questionnaire and printouts. In the manual system, correction also included writing the correction on the printout, which was then passed to someone else to make the actual changes on line.

- During census collection it is possible to receive more than one questionnaire for a given farming operation. Procedures are carried out to identify and remove this duplication. Access to the images of the potential duplicates using split screens resulted in significant savings in time.

- An undercoverage problem during collection resulted in a need to identify farms, which were missed. To identify them matches between the 1996 base, tax, surveys and previous census lists was carried out. Access to the images allowed editors to make a last quick check if the farm was on the base. Field follow-up costs, response burden and the probability of creating duplicate questionnaires were all reduced. Images provided information (e.g. land descriptions etc.) that were not on captured database.

- Hogs grown under contract for others is a rapidly growing service in Canadian agriculture. Many farmers growing hogs under contract failed to report the hogs on their questionnaires in spite of instructions to do so. The hogs didn't belong to them so they didn't include them. We had the farm but not the hogs. Matches between the 1996 census and the semi-annual hog survey identified farms reporting large numbers of hogs on the survey but not on the census. Were the hogs missing due to processing or respondent error? A quick check of the images answered the question. Follow-up procedures were then carried out with the respondents to add the missing hogs.

- The automated imputation procedures always seemed to leave a few situations for each module of questions where a donor can not be found. In many cases the cause is respondent or processing errors. With images subject matter experts quickly corrected these situations. In previous censuses "uninformed quick fixes" were made just to move the questionnaires to the next imputation stage. The time to find and pull the physical documents would have seriously delayed the imputation process.

Did imaging pay for itself in one cycle?

No and nor should it have been expected to. How do you put an economic value on quality of data, job satisfaction, or the cost of a delay in the release date?

Will imaging pay for itself in the short term (5 years)? Without question yes. The equipment continues to be used by the census for testing, plus on many other projects on testing and production. The equipment will also, along with upgrades and additions, play a role in the 2001 census. Census of Agriculture's positive experience has encouraged others to test and implement not only imaging but ICR applications. Areas now involved in the technology include:

- Whole Farm Data Project
- Census of Population
- Tax Data Division
- Industrial Organization and Finance Division
- Operations and Integration Division
- Administrative Support Services Division (Records Management)
- Project to Improvement Provincial Economic Statistics (PIPES)

In some cases the Census of Agriculture has taken on work for other projects. This additional work:

- helps pay for the system, plus share the maintenance and licensing costs
- supports related research
- expands/maintains experience of the staff
- adds more features and capabilities to the system

The decision to implement imaging was unquestionably correct. The level of risk taken in implementing was acceptable but not insignificant. Any additional challenges would have put the entire project at risk.

Experience has proven that the risk associated with implementing imaging was well worth taking. In the end it played a vital role in aiding the project to meet its planned release dates.

A number of events pulled the project off schedule. The undercoverage problems alone had put the data validation task several weeks behind schedule. In addition problems with the generalized imputation system forced the development of a major fix and re-imputation of all the financial variables for two of the largest provinces.

A total team effort, well above and beyond the call of duty, plus efficiencies generated from access to the images allowed the release of the data base to take place on schedule.

Where does technology take the 2001 Census of Agriculture? Can ICR handle 16 page legal sized forms completed by 300,000 aging farmers?

Once again there are different opinions and camps. Some project members who were in the conservative camp in 1996 are in the progressive camp now and vice versa. A number of other things have changed. We know imaging works and works well, we have more experience, confidence and we have more time.

ICR research for the Census of Agriculture started almost one year ago, four years before the 2001 census. Early results are encouraging but it is far too early for any predictions let alone decisions. An assessment of respondent reaction to ICR friendly questionnaires will begin in 1998.

What is the fall back position if ICR does not prove to be cost effective or appropriate for the project? Imaging with more parallel processing is a conservative option. Mark Character Recognition (MCR) combined with key-from-image would be another significant leap forward.

Document imaging is considered by some to be a transient technology as the move to the paperless office continues. A question to ponder is, «When will electronic reporting replace paper questionnaires for the Census of Agriculture?»

RELATED DOCUMENTS

1. Bradshaw, C. & Duggan, J. (1997) A Review of Document Imaging for the 1996 Census of Agriculture.
2. Duggan, J. (1996) Electronic Imaging in Support of the 1996 Census of Agriculture. Presented at the International Conference on Computer-Assisted Survey Information Collection.
3. Green, I. (1995) Questionnaire Imaging Pilot. Internal Report. Statistics Canada.

Agriculture and Rural Working Papers Series

The Agriculture Division publishes working papers on research, analytical results, statistical techniques, methods and concepts. Contact the Agriculture Division by calling toll-free 1-800-465-1991 to order your copy. (Some of the Working Papers are available on the Statistics Canada Internet site).

No. 1	(21-601-MPE80001)	**A Description of Theil's RMPSE Method in Agricultural Statistical Forecasts (1980)**	Stuart Pursey
No. 3	(21-601-MPE81003)	**A Review of the Livestock Estimating Project with Recommendations for the Future (1981)**	Bernard Rosien and Elizabeth Leckie
No. 4	(21-601-MPE84004)	**An Overview of the Canadian Oilseed Industry (1984)**	Glenn Lennox
No. 5	(21-601-MPE84005)	**Preliminary Analysis of the Contribution of Direct Government Payments to Realized Net Farm Income (1984)**	Lambert Gauthier
No. 6	(21-601-MPE84006)	**Characteristics of Farm Entrants and their Enterprises in Southern Ontario for the Years 1966 to 1976 (1984)**	Jean B. Down
No. 7	(21-601-MPE84007)	**A Summary of Commodity Programs in the United States (1984)**	Allister Hickson
No. 8	(21-601-MPE84008)	**Prairie Summerfallow Intensity: An Analysis of 1981 Census Data (1984)**	Les Macartney
No. 9	(21-601-MPE85009)	**The Changing Profile of the Canadian Pig Sector (1985)**	Mike Shumsky
No. 10	(21-601-MPE86010)	**Revisions to the Treatment of Imputed House Rents in the Canadian Farm Accounts, 1926 - 1979 (1986)**	Mike Trant
No. 11	(21-601-MPE92011)	**The Ratio Estimator: an Intuitive Explanation and Its Use in Estimating Agriculture Variables (1992)**	François Maranda and Stuart Pursey
No. 12	(21-601-MPE91012)	**The Impact of Geographic Distortion Due to the Headquarters Rule (1991)**	Rick Burroughs
No. 13	(21-601-MPE91013)	**The Quality of Agriculture Data – Strengths and Weaknesses (1991)**	Stuart Pursey
No. 14	(21-601-MPE92014)	**Alternative Frameworks for Rural Data (1992)**	A.M. Fuller, Derek Cook and Dr. John Fitzsimons
No. 15	(21-601-MPE93015)	**Trends and Characteristics of Rural and Small Town Canada (1993)**	Brian Biggs, Ray Bollman and Michael McNames
No. 16	(21-601-MPE92016)	**The Microdynamics and Farm Family Economics of Structural Change in Agriculture (1992)**	Phil Ehrensaft and Ray Bollman
No. 17	(21-601-MPE94017)	**Grains and Oilseeds Consumption by Livestock and Poultry, Canada and Provinces 1992**	Livestock and Animal Products Section
No. 18	(21-601-MPE94018)	**Trends and Patterns of Agricultural Structural Change: Canada / US Comparison**	Ray Bollman, Leslie A. Whitener and Fu Lai Tung
No. 19	(21-601-MPE94019)	**Farm Family Total Income by Farm Type, Region and Size for 1990 (1994)**	Saiyed Rizvi, David Culver, Lina Di Piétro and Kim O'Connor
No. 20	(21-601-MPE94020)	**Adjustment in Canadian Agriculture (1994)**	George McLaughlin
No. 21	(21-601-MPE93021)	**Microdynamics of Farm Size Growth and Decline: A Canada-United States Comparison**	Fred Gale and Stuart Pursey

Statistics Canada Statistique Canada

Canada

No. 22	(21-601-MPE92022)	**The Structures of Agricultural Household Earnings in North America: Positioning for Trade Liberalization -**	Leonard Apedaile, Charles Barnard, Ray Bollman and Blaine Calkins
No. 23	(21-601-MPE92023)	**Potatoes: A Comparison of Canada/USA Structure**	Glenn Zepp, Charles Plummer and Barbara McLaughlin
No. 24	(21-601-MPE94024)	**Farm Structure Data: A US-Canadian Comparative Review**	Victor J. Oliveira, Leslie A. Whitener and Ray Bollman
No. 25	(21-601-MPE94025)	**Grain Marketing Statistics Statistical Methods Working Paper Version 2**	Karen Gray
No. 26	(21-601-MPE94026)	**Farm Business Performance: Estimates from the Whole Farm Database**	W. Steven Danford
No. 27	(21-601-MPE94027)	**An Attempt to Measure Rural Tourism Employment**	Brian Biggs
No. 28	(21-601-MPE95028)	**Delineation of the Canadian Agricultural Ecumene for 1991**	Timothy J. Wershler
No. 29	(21-601-MPE95029)	**Mapping the Diversity of Rural Economies: A Preliminary Typology of Rural Canada**	Liz Hawkins
No. 30	(21-601-MPE96030)	**Structure and Trends of Rural Employment: Canada in the Context of OECD Countries**	Ron Cunningham and Ray D. Bollman
No. 31	(21-601-MPE96031)	**A New Approach to Non-CMA\CA Areas**	Linda Howatson-Leo and Louise Earl
No. 32	(21-601-MPE96032)	**Employment in Agriculture and Closely Related Industries in Rural Areas: Structure and Change 1981-1991**	Sylvain Cloutier
No. 33	(21-601-MPE98033)	**Hobby Farming – For Pleasure or Profit?**	Stephen Boyd
No. 34	(21-601-MPE98034)	**Utilization of Document Imaging Technology By the 1996 Canadian Census of Agriculture**	Mel Jones and Ivan Green
No. 35	(21-601-MPE98035)	**Employment patterns in the non-metro workforce**	Robert Mendelson
No. 36	(21-601-MPE98036)	**Rural and small town population is growing in the 1990s**	Robert Mendelson and Ray D. Bollman
No. 37	(21-601-MPE98037)	**The composition of business establishments in smaller and larger communities in Canada**	Robert Mendelson
No. 38	(21-601-MPE98038)	**Off-farm Work by Census-farm Operators: An Overview of Structure and Mobility Patterns**	Michael Swidinsky, Wayne Howard and Alfons Weersink
No. 39	(21-601-MPE99039)	**Human Capital and Rural Development: What Are the Linkages?**	Ray D. Bollman
No. 40	(21-601-MPE99040)	**Computer Use and Internet Use by Members of Rural Households**	Margaret Thompson-James
No. 41	(21-601-MPE99041)	**RRSP Contributions By Farm Producers in 1994**	Marco Morin
No. 42	(21-601-MPE99042)	**Integration of Administrative Data with Survey and Census Data**	Michael Trant and Patricia Whitridge

Prices:

1 copy	**Free**
2 to 10 copies	**$50.00**
11 to 25 copies	**$75.00**
26 to 38 copies	**$100.00**

ORDER FORM

Statistics Canada

TO ORDER:

MAIL
Agriculture Division
Statistics Canada
12th floor, Jean Talon Bldg.
Tunney's Pasture
Ottawa, Ontario
K1A 0T6

PHONE
1 800 465-1991
Charge to VISA or MasterCard. Outside Canada and the U.S., and in the Ottawa area call (613) 951-7134. Please do not send confirmation.

FAX
1 613 951-1680
VISA, MasterCard and purchase orders only. Please do not send confirmation. A fax will be treated as an original order.

INTERNET agriculture@statcan.ca

1 800 363-7629
Telecommunication Device for the hearing Impaired

(Please Print)

Company _____ Department _____

Attention _____ Title _____

Address _____

City _____ Province _____

( ) _____ ( ) _____

Postal Code Phone Fax

E-mail address: _____

METHOD OF PAYMENT:

(Check only one)

☐ **Please Charge my:** ☐ VISA ☐ MasterCard

Card Number _____

Expiry Date _____

Cardholder (*please print*) _____

Signature _____

☐ **Payment enclosed $** _____

☐ **Purchase Order Number** _____
(please enclose)

Authorized Signature _____

Working Paper Number	Title	Catalogue Number	No. of copies	Total $

PRICES:

1 copy	Free
2 to 10 copies	$50.00
11 to 25 copies	$75.00
26 to 38 copies	$100.00

- **Prices are subject to change. To confirm current prices call 1 800 465-1991.**
- **Catalogue prices for clients outside Canada are shown in US dollars. Clients outside Canada pay total amount in US funds drawn on a US bank.**
- **Canadian clients pay in Canadian funds and add 7% GST and applicable PST or HST.**
- **Cheque or money order should be made payable to the *Receiver General for Canada.***
- **GST Registration # R121491807**

SUBTOTAL (please circle Canadian or U.S. funds)	**Cnd./ U.S. $**	**$**
Discount (if applicable)		
GST (7%) (Canadian clients only, where applicable)		
Applicable PST (Canadian clients only, where applicable)		
Applicable HST (N.S., N.B., Nfld.)		
GRAND TOTAL		**$**

PF097177

THANK YOU FOR YOUR ORDER!

Statistics Canada Statistique Canada

Canada

Statistique Canada
Division de l'agriculture

DOCUMENT DE TRAVAIL #34

Utilisation de la technologie d'imagerie documentaire dans le recensement canadien de l'agriculture de 1996

Mel Jones et Ivan Green
Statistique Canada

Février 1998

N° 21-601-MIF98034 au catalogue

L'analyse et l'interprétation des données sont la responsabilité de l'auteur et non celle de Statistique Canada.

© Ministre de l'Industrie, Statistique Canada, 1998. Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu du présent document, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement.

Utilisation de la technologie d'imagerie documentaire

dans le recensement canadien de l'agriculture de 1996

Mel Jones et Ivan Green
Statistique Canada
le 3 février 1998

Sommaire

Le traitement des données du recensement canadien de l'agriculture dépend grandement de l'accès à ses questionnaires. On a besoin de ces documents à bien des stades de la collecte et du traitement pour que des produits de qualité puissent être livrés à la clientèle agricole. L'accès physique aux documents s'est révélé difficile par le passé sur le plan de la remise des questionnaires en temps utile, de l'accès simultané, de la gestion, du repérage, de l'entreposage et des coûts. Dès 1984, les responsables du recensement ont reconnu que la technologie de l'imagerie documentaire pouvait les aider à améliorer la qualité et les délais de livraison des données. Il reste que, dans les années 1980, on se rendait compte d'emblée que la technologie du moment n'était pas assez mûre ni abordable pour le recensement. En 1996, l'imagerie documentaire est devenue une fructueuse réalité pour le recensement canadien de l'agriculture. Dans ce document, nous décrirons les expériences, les enjeux et les attentes des intervenants très divers qui ont été associés à l'implantation de cette technologie.

Introduction

L'implantation de la technologie d'imagerie documentaire aux fins du recensement canadien de l'agriculture a comporté des frais et des risques de taille. Dès les premiers stades du projet, on a amplement discuté des méthodes classiques et des technologies rivales. Dans une première section, nous exposerons les questions débattues et dégagerons les orientations retenues. Nous mettrons en parallèle les inconvénients des méthodes manuelles du passé et les avantages attendus de cette technologie.

Nombreuses ont été les difficultés liées à l'implantation d'un système d'imagerie documentaire. Dans une deuxième section, nous évoquerons plusieurs des défis relevés et dirons comment les grands problèmes techniques qui se posaient ont été cernés et résolus. Suivront des observations au sujet des raisons, évidentes ou non, pour lesquelles ce projet d'implantation a été couronné de succès.

À la troisième section, nous parlerons de l'optique de la direction dans la décision prise de recourir à l'imagerie, ainsi que des possibilités de maintien et d'évolution de cette technologie à Statistique Canada.

Section I : Pourquoi recourir à l'imagerie?

Selon la politique adoptée par l'organisme, les dates de diffusion des publications sont communiquées d'avance au public. Une fois déclarées, ces dates ne changent plus, aussi la production statistique est-elle assujettie à des échéances très fermes en fonction des dates de diffusion.

Les contraintes de temps et de budget, jointes à la nécessité de maintenir, voire d'améliorer la qualité, ont suscité un mouvement d'optimisation de l'efficience. C'est pourquoi on mise davantage sur l'automatisation et l'innovation. Sont principalement visées ici les méthodes manuelles coûteuses.

Pour le recensement canadien de l'agriculture, on s'est surtout attaché a) à la saisie des données et b) à la remise des questionnaires tout au long des activités de production. On a réalisé des études de faisabilité de l'automatisation pour l'un et l'autre de ces aspects.

En ce qui concerne la saisie des données, on a étudié la technologie de reconnaissance intelligente de caractères (RIC). À l'époque, on s'interrogeait sur le mûrissement de la technologie RIC aux fins du recensement, et on manquait de temps pour vérifier les réactions des recensés. Si on avait utilisé cette

technologie, on aurait nettement alourdi le fardeau de réponse, car il aurait fallu passer de réponses numériques librement consignées par les recensés à des réponses numériques assujetties à des contraintes et adaptées à la technologie RIC. Comme plus de 90 % des renseignements sollicités dans le cadre du recensement sont numériques, on aurait sans doute eu l'obligation de repenser le questionnaire de fond en comble. Ajoutons que, après l'étude de faisabilité, on s'est posé des questions sur l'efficacité et la rentabilité des logiciels de reconnaissance de caractères.

La recherche sur l'imagerie documentaire et la technologie RIC a créé une situation très intéressante et riche en défis. Pour emprunter des termes de politique canadienne, disons que les membres de l'équipe de gestion du projet se sont divisés en trois camps, ceux des conservateurs, des progressistes-conservateurs et des progressistes. Les conservateurs jugeaient que l'implantation en soi d'un système d'imagerie était trop risquée pour les raisons suivantes :

- Le recensement allait être tenu dans deux ans et, dans une telle entreprise quinquennale, il était trop hasardeux de se mettre à essayer et à implanter un tel système en deux ans.
- Selon les prévisions, les coûts varieraient amplement d'abordables à franchement inabordables.
- L'organisme ne connaissait rien de l'imagerie documentaire ni de la technologie RIC, aussi était-ce user de prudence que d'éprouver ces techniques dans des enquêtes plus modestes et plus fréquentes.
- Le personnel dépendrait entièrement des machines. Si le système tombait en panne, le personnel n'avait plus rien à faire.
- Le projet offrait de nombreux autres défis, dont les suivants :
    - On avait ajouté une nouvelle enquête sur l'état d'avancement des semailles. C'était là une enquête de suivi ITAO qui s'adressait à plus de 40 % des recensés.
    - Il fallait entièrement repenser les méthodes et passer d'une exploitation en ordinateur central à l'utilisation d'ordinateurs intermédiaires.
    - Il fallait un système et une méthodologie nouveaux pour le Registre des fermes.

Les conservateurs voyaient bien l'intérêt et les avantages de l'imagerie, mais en jugeaient les risques trop grands pour le recensement de l'agriculture de 1996. La méthode manuelle du passé (sans imagerie) était éprouvée et la plus sûre.

Aux yeux des progressistes, on économiserait le plus en implantant conjointement un système d'imagerie et la technologie RIC. Sans cette technologie, le système d'imagerie ne serait pas rentable. Les tenants de ces vues étaient persuadés non seulement qu'un système d'imagerie était réalisable, mais aussi que la technologie RIC avait alors suffisamment mûri pour les besoins du recensement.

Le chargé de projet avait d'abord proposé d'implanter le système d'imagerie sans RIC. Les progressistes-conservateurs l'avaient appuyé. Ils jugeaient inacceptables les risques d'une implantation RIC dans les délais à respecter. Ils pensaient par ailleurs qu'un système d'imagerie était faisable sans trop de risques. On leur avait parlé d'une foule de projets d'imagerie qui avaient tourné à la catastrophe, mais ils avaient aussi pu observer assez d'applications qui fonctionnaient ailleurs pour acquérir la conviction que le jeu en valait la chandelle.

Dans ce débat, aucun camp n'avait tort ou raison, chacun ayant de bonnes raisons de prendre la position qu'il avait adoptée. Une fois la décision prise d'implanter un système d'imagerie sans technologie RIC, tous les membres de l'équipe de projet se sont totalement rangés à cet avis.

Le but était maintenant de communiquer des images électroniques des questionnaires aux ordinateurs de bureau des vérificateurs et des valideurs.

Si on avait exploité un système manuel de remise de questionnaires pour 1996, celui-ci aurait accusé les insuffisances de ceux qui l'avaient précédé :

- ***Livraison en temps utile.*** Il faudrait compter des heures, voire des jours pour la remise des questionnaires au personnel de production.
- ***Accès simultané.*** C'est là une autre question de livraison en temps utile. Souvent, les valideurs et les vérificateurs doivent avoir simultanément accès aux questionnaires de grandes exploitations agricoles. Ce sont ces dernières qui tiennent la plus grande place dans les estimations de nombreuses variables du recensement. La livraison serait successive, car on ne pourrait tirer des copies des questionnaires, activité refusée du fait de ses risques pour la sécurité. Pour contrôler les questionnaires, on les archivait et les désarchivait à chaque opération ou à chaque demande.

- ***Repérage.*** Pour contrôler le cheminement individuel des documents, il faudrait se doter d'un système de repérage automatisé. À chaque stade de la production, on aurait besoin d'accuser réception manuellement après un passage en lecteur de code à barres. Un système de ce genre imposerait un cheminement successif des questionnaires (et des données) dans le circuit.
- ***Archivage complexe.*** On aurait aussi besoin de tout un système de classement par chemises, boîtes et étagères pour un archivage et un désarchivage à grand volume.
- ***Coûts élevés.*** Avec les préposés, les étagères, les locaux d'archivage, les mesures de sécurité et la mise au point d'un système de distribution et de repérage de questionnaires, on se retrouverait avec un système manuel coûteux et relativement lent.
- ***Frustration.*** Les valideurs devaient dresser sur papier la liste des questionnaires dont ils avaient besoin et la présenter à la bibliothèque. En temps normal, on répondrait à leurs demandes en quelques heures, mais beaucoup de questionnaires manqueraient à cause de l'accès simultané, d'erreurs de classement, etc. Il faudrait produire plusieurs demandes avant que tous les documents désirés ne soient réunis.
- ***Erreurs de classement.*** Des erreurs de classement se produisent lorsqu'on archive et désarchive. Il faudrait interrompre périodiquement les opérations en bibliothèque pour effectuer une vérification systématique et exhaustive.

Par intuition, on pouvait voir qu'un système d'imagerie de questionnaires permettait de remédier aux insuffisances d'un système manuel :

- ***Livraison en temps utile.*** Il faudrait non plus des heures ou des jours mais quelques secondes pour communiquer des images de questionnaires au personnel de production.
- ***Accès simultané.*** Tous les utilisateurs pourraient examiner simultanément un même document, d'où des possibilités de traitement en ***parallèle***.
- ***Repérage.*** Tout au long du traitement, on n'avait plus besoin de système manuel de repérage de questionnaires. Les gens n'avaient pas non plus à inscrire les entrées et les sorties de questionnaires.
- ***Archivage complexe.*** On aurait encore besoin d'un système temporaire d'archivage, mais il serait relativement plus simple.
- ***Coûts élevés.*** Comme on n'avait jamais fait d'imagerie à grande échelle à Statistique Canada, on n'avait pas de point de comparaison pour le recensement de l'agriculture, aussi ne savait-on au juste si un système d'imagerie coûtait plus cher ou moins cher qu'un système manuel de remise de questionnaires.

La direction a jugé que les avantages possibles méritaient un examen plus attentif.

Outre les avantages énumérés, elle prévoyait les retombées suivantes :

- on n'aurait pas besoin de microfilmer les documents après traitement.
- on réduirait les opérations manuelles, notamment celles des préposés à la bibliothèque.
- grâce à un accès immédiat aux documents, le personnel gagnerait en efficience, car il pourrait constater les problèmes et les résoudre immédiatement; on n'aurait plus besoin de dresser la liste des problèmes et de trouver autre chose à faire jusqu'à ce que les documents arrivent.
- les commis et les valideurs auraient tous moins de papier à manipuler; la paperasserie serait en grande partie éliminée par la communication d'images électroniques.
- on n'aurait plus la frustration que causent l'attente de questionnaires et les documents en retard ou manquants.
- enrichissement des tâches : le personnel jugerait désormais satisfaisantes et stimulantes les tâches fastidieuses du passé.
- en ayant un accès immédiat aux questionnaires problèmes, les préposés compteraient plus sur l'information reçue que sur leur intuition; pour ne pas avoir à attendre de documents sur papier, ils auraient parfois tendance à deviner les problèmes et les solutions.
- grâce à un accès immédiat, le personnel pourrait considérer plus de problèmes et ainsi améliorer la qualité des estimations, notamment à de bas niveaux de description géographique.

Section II : Questions d'imagerie

Lorsqu'on s'est mis à étudier l'imagerie documentaire, on a voulu répondre aux questions suivantes :

- Qu'est-ce que l'imagerie documentaire?
- Quels éléments d'imagerie permettront-ils de combler nos besoins?
- Cette imagerie fonctionnera-t-elle dans notre cadre technique?

- À quoi pourraient se monter les coûts en gros?
- Quelles questions doit-on étudier si on veut que l'implantation soit réussie?

Les représentants du secteur du recensement ont conféré en première étape avec le secteur privé. Ils ont d'abord communiqué à l'industrie les critères, les volumes et les spécimens du questionnaire de 1991 qu'ils jugeaient représentatifs du questionnaire de 1996 (qui n'avait pas encore été élaboré). Les représentants de l'industrie sont revenus avec des prévisions de coûts et des solutions proposées.

Selon leurs estimations, il fallait prévoir de 500 000 à presque 3 000 000 dollars canadiens. Le premier chiffre était acceptable et le second non.

Les propositions de solution prêtaient quelque peu à confusion. On y abordait des problèmes qui n'avaient nettement rien à voir avec le recensement et en oubliait d'autres qui risquaient de se poser. Il reste que, malgré un peu de confusion, les représentants du secteur du recensement ont commencé à se faire une meilleure idée des enjeux.

Voici notamment des questions qui semblaient se poser (et dont la solution serait fort coûteuse) pour le recensement :

- indexation;
- contrôle de la qualité;
- débit des lecteurs automatiques sur papier ministre Vista 70M (questionnaire du recensement);
- modifications à prévoir à la disposition des questionnaires pour les adapter au ***système d'imagerie***;
- qualité des images.

Pour pouvoir dresser une liste suffisante de questions, on a assisté à des séminaires sur l'imagerie documentaire dans le cadre des salons spécialisés de l'Association for Information and Image Management (AIIM) et rendu visite à des chercheurs en imagerie du United States Bureau of the Census (USBC).

Les représentants de l'USBC ont mis en lumière la question primordiale de l'***intégrité des documents***. Les questionnaires de cet organisme se présentent sous forme de livret (plusieurs pages). Les lecteurs traitent un feuillet à la fois de sorte que, pour capter un document de plusieurs pages, on peut être obligé de démembrer un questionnaire en autant de pages. Ce qui inquiétait avant tout les chercheurs de l'USBC, c'était le risque que des pages de tels documents s'égarent ou ne soient plus dans le bon ordre avant ou après la lecture, c'est-à-dire le risque de perte d'intégrité des documents. À l'époque, l'USBC étudiait des dispositifs tourne-pages (plutôt que coupe-pages). Comme le questionnaire du recensement de l'agriculture comprend plusieurs pages, on a vu que le problème se posait aussi à Statistique Canada. Ces questionnaires étant relativement peu nombreux (280 000), les dispositifs tourne-pages ne paraissaient pas une solution rentable.

Une des premières solutions envisagées a été d'apposer un code unique à chaque page d'un questionnaire. On s'est entretenu à ce sujet avec des spécialistes internes et des représentants de grandes imprimeries. La solution était certes réalisable, mais son rapport efficacité-coût était inacceptable, puisqu'on devait prévoir près de 400 000 dollars canadiens, ajoutant ainsi 1,43 $ environ au coût d'impression de chaque questionnaire, ce qui ferait plus que tripler le budget d'impression. Avec cette solution, il y avait toujours des risques d'erreur de mise en ordre. Comme autre solution, on a songé à faire imprimer en relief un numéro d'ordre au bas de chaque page par lecteur automatique. C'était là la bonne solution pour les questionnaires égarés ***après*** lecture, mais non avant lecture. Il fallait donc pousser l'étude de la question.

On a continué à examiner la question de l'imagerie documentaire. Les salons spécialisés nous ont éclairés sur d'autres questions :

- taille et qualité des moniteurs;
- cartes vidéo et fonctionnalité;
- capacité réseau;
- capacité de stockage;
- fonctionnalité des lecteurs;
- soutien technique;
- intégration aux applications internes.

On a fini par dresser une liste relativement exhaustive, mais on ne savait encore au juste dans quelle mesure ses éléments s'appliquaient tous au recensement de l'agriculture. On craignait que l'application du recensement comporte des questions particulières d'imagerie à régler. Ainsi, la plupart des lecteurs sont conçus pour traiter des feuillets uniques format lettre de force 30. Pour le recensement, on utilise des livrets de 16 pages de papier ministre Vista 70M. La direction a accepté que l'on réalise un petit projet pilote (au coût de 20 000 dollars canadiens) pour clarifier les questions relatives au recensement.

On a demandé à une entreprise locale d'implanter à contrat un système d'imagerie de base sur place. Elle a travaillé en étroite collaboration avec les responsables des systèmes internes et les spécialistes des méthodes. L'exercice visait :

- à étudier les méthodes et les procédures d'exécution nécessaires à l'implantation d'un système d'imagerie entièrement fonctionnel pour ce qui est de la réception, de la préparation, de la lecture, du remontage, de l'archivage, du désarchivage et du contrôle de la qualité des questionnaires;
- à apprendre à connaître le matériel et les logiciels types d'imagerie de l'industrie;
- à se renseigner sur le rendement de cette technologie dans le cas des documents du recensement.

Le système pilote se composait des éléments suivants :

- deux lecteurs (Fujitsu et Bell + Howell);
- un serveur Hewlett-Packard sur UNIX et LAN-Manager 2.1;
- quatre PC dotés de cartes vidéo de 2 Mo;
- moniteurs couleur et échelle de gris de 17 et 21 po;
- TCP/IP;
- sous-réseau Ethernet 10 Base T UTP;
- visualisateurs d'imagerie Watermark TIF;
- échantillon de 2 000 questionnaires de 16 pages papier ministre du recensement de 1991.

Le projet pilote a duré trois jours et donné les résultats suivants :

- La technologie a été «politiquement» mieux acceptée. L'observation des questionnaires en lecture, de la qualité des images et de la fonctionnalité des visualisateurs a accru la confiance des gens et atténué les réactions de scepticisme.
- Comme Statistique Canada s'était conformé aux normes d'interface, il n'y a pas eu de surprises tenant à l'infrastructure technique.
- Le débit observé des lecteurs était considérablement moindre que prévu, mais impressionnait quand même.
- Dans cette expérience faite dans le secteur du recensement, l'équivalence de modèles de lecteurs bien entretenus venant de fabricants concurrents n'est pas synonyme d'équivalence de rendement. Un des lecteurs a été d'un fonctionnement sûr avec relativement peu de captation en double, d'images obliques et de blocage. L'autre a été d'un rendement fort médiocre. Comme il détruisait littéralement les questionnaires, on l'a rapidement retiré du projet pilote.
- La fonctionnalité des lecteurs était très coûteuse (lecture de codes à barres, indexation, comptage, etc.).
- On a pu établir le nombre de lecteurs de ce type et le nombre de préposés dont on avait besoin pour les quantités de questionnaires de recensement à traiter.
- La qualité des images était acceptable (200 ppp).
- On a déterminé la taille moyenne des images (produites par les lecteurs du projet pilote); avec ces indications, on pourrait mieux simuler le trafic réseau et établir les besoins de stockage d'images.
- La force et la couleur du papier des questionnaires (papier Vista opaque 70M) étaient acceptables; il n'y avait pas de «lavage» de caractères d'une extrémité à l'autre d'un feuillet.
- La disposition du questionnaire de 1991 ne se prêtait pas à l'imagerie, mais les corrections à apporter seraient légères et n'auraient aucune incidence sur les recensés ni sur le budget d'impression en 1996. On n'avait pas à vérifier la réaction des recensés à un questionnaire adapté à l'imagerie.
- La fonctionnalité, l'aspect et l'impression qu'offre un visualisateur Watermark seraient conformes aux exigences du recensement, ce qui permettrait d'estimer grossièrement les coûts afférents et d'obtenir des valeurs de référence à des fins de spécification technique.
- Le problème d'intégrité des documents ne se posait pas. On avait craint que, si des documents coupés tombent, des pages soient égarées ou ne soient plus dans le bon ordre, d'où une perte d'intégrité. Des essais ont démontré que les liasses de questionnaires coupés ne se défaisaient généralement pas si elles tombaient à cause de l'électricité statique et des plis des pages. Le risque que des pages se perdent ou s'égarent n'en subsistait pas moins. On a pris des mesures pour l'atténuer encore plus. On a notamment gardé les liasses petites (30 documents au plus), les a rangées dans des enveloppes en réduisant au minimum le temps passé hors pli, diminué la distance entre le massicot et le lecteur, fait le remontage des documents avec des caractères manuscrits et des couleurs d'encre caractéristiques et, comme dernière précaution, consulté les images ou les données de la base de données structurées en vue du remontage des documents qui étaient tombés.

- Des moniteurs couleur 17 po moins chers avec leurs cartes vidéo étaient acceptables (résolution 1 024x768) pour au moins la moitié des activités de production du recensement. L'image était souvent d'une qualité ***supérieure*** à celle de l'original, parce que les images pouvaient être grossies. Avec les moniteurs à échelle de gris, l'affichage était plus clair, mais pas outre mesure. On jugeait en outre plus facile d'affecter les moniteurs couleur à d'autres activités après la production.

Les résultats du projet pilote ont frayé la voie à l'implantation du système. Voici ce qu'il restait à faire :

- élaboration d'une stratégie de contrôle de qualité;
- élaboration d'une stratégie d'indexation;
- établissement des besoins réseau;
- établissement des besoins en intégration;
- établissement des besoins en lecteurs;
- établissement des besoins en visualisateurs;
- choix et acquisition de moniteurs et de cartes vidéo;
- acquisition d'une technologie de stockage;
- conception de système d'exécution et de contrôle de lecture;
- établissement d'une demande de propositions (DDP);
- implantation.

Dans tout le reste du projet, les contributions respectivement apportées par les représentants de la Division des services aux utilisateurs de l'informatique (DSUI), du Groupe de soutien aux ordinateurs de gamme moyenne, de la Division des méthodes d'enquêtes-entreprises et de la Division du développement des systèmes de Statistique Canada ont été primordiales.

Largeur de bande : En ce qui concerne les charges de production en imagerie, un grand sujet d'inquiétude était le réseau. Quelques années avant le projet, l'organisme s'était recâblé en 10 Base T UTP. Le nouveau câblage s'était révélé vulnérable lorsque de grands fichiers étaient transmis. On devait s'assurer qu'il pouvait soutenir les activités simultanées de 100 utilisateurs finals et des opérations de lecture à grand volume. L'absence de point de comparaison dans l'organisme pour les charges de traitement d'images posait un problème. Heureusement, on pouvait toujours consulter les principaux responsables de l'exécution du recensement précédent et s'enquérir des quantités de questionnaires qui étaient quotidiennement passées par le système manuel. On prévoyait que les volumes d'images dépasseraient ceux des documents sur papier. Pour vérifier, on a triplé les quantités traitées comme trafic utilisateur. On a ajouté cette quantité au volume prévu de trafic lecteur (nombre total de questionnaires divisé par le nombre total de jours disponibles de lecture) pour représenter la charge quotidienne totale du réseau en imagerie. On a divisé ce total par 6,5 heures (nombre moyen d'heures d'activité dans une journée de 7,5 heures) pour obtenir la charge moyenne en cours de production. À l'aide de cette formule et des données sur la taille des images fournies par le projet pilote, on a mené divers essais réseau.

On a sollicité l'aide des spécialistes du Centre de services en communication (CSC). Ceux-ci ont indiqué comment s'y prendre efficacement pour réaliser des essais qui donneront les résultats les plus utiles. Pour représenter l'activité de 100 utilisateurs, il ne suffisait pas de simuler la demande en mégaoctets-seconde, c'est-à-dire le fonctionnement d'un poste de travail en communication d'imagerie dans les deux sens avec un serveur dans le réseau. Il importait aussi de simuler la répartition de la demande. Si un seul poste faisait venir des images et un autre en envoyait, des collisions étaient improbables en réseau. Il fallait donc se servir d'un grand nombre de postes pour que les essais soient significatifs.

Avant de procéder à ces essais, le CSC a configuré un sous-réseau isolé. Il a éliminé le bruit pour que les charges mesurées soient attribuables à la transmission d'images. On a fait marcher 20 postes de travail avec chacun un programme qui téléchargeait un fichier en provenance ou à destination d'un serveur dans un laps de temps donné. Le CSC a pris des analyseurs de réseau appelés «renifleurs» pour observer le trafic.

Il a recommandé l'acquisition d'un autocommutateur à routage Ethernet d'une capacité de 100 Mo/s pour les charges prévues en production. Des représentants du Groupe de soutien technique ont installé l'autocommutateur et isolé les utilisateurs du recensement du reste de l'organisme. Le but était double : (1) protéger le sous-réseau d'imagerie contre toute dégradation du rendement imputable à des variations imprévues du trafic de réseaux locaux extérieurs et (2) protéger ces réseaux contre les charges d'images du secteur du recensement.

On a réparti les utilisateurs du recensement en trois groupes de travail dans l'autocommutateur Ethernet avec une liaison de 10 Mo/s. On a relié le serveur à

l'autocommutateur par un câble optique de 100 Mo/s. Le secteur du recensement a payé une carte FDDI (Interface optoélectronique de données à mode réparti) pour l'interface serveur-réseau. Le groupe de soutien aux ordinateurs de gamme moyenne a étudié les solutions possibles, présenté une recommandation et, après consultation, fait l'acquisition et l'installation de la carte FDDI.

Après l'implantation du *sous-réseau d'imagerie*, le CSC a procédé à de nouveaux essais. Il a ainsi pu vérifier que les besoins avaient bel et bien été comblés.

Moniteurs : Comme le système d'imagerie est au service de l'utilisateur final et qu'il n'est souvent connu que par ce qui paraît à l'écran, le moniteur en est sans doute l'élément primordial, et sûrement un des plus coûteux. Pendant le projet pilote, on n'a relevé aucune différence importante entre le moniteur couleur de 17 po et le moniteur échelle de gris de 19 po pour la qualité de perception des images du recensement. Comme les moniteurs devaient être utilisés dans plusieurs applications d'imagerie, on a décidé de faire l'acquisition de moniteurs couleur malgré leur prix supérieur. Il restait à établir a) la (les) taille(s) d'écran convenant aux besoins du recensement et b) les exigences relatives aux cartes vidéo.

À certains stades des activités de production, on doit prévoir l'observation simultanée de deux pages format ministre ou plus. On ignorait si un moniteur de 17 po avec une carte vidéo bien configurée suffirait ou non. L'écart de prix est très appréciable entre les écrans de 17 et de 21 po. Il s'agissait de voir si on avait effectivement besoin de grands moniteurs chers.

Le personnel du recensement disposait alors de plusieurs moniteurs de 17 po pouvant servir à une évaluation en imagerie, mais il n'y avait pas de moniteurs plus grands. On a pris contact avec des représentants de l'entreprise privée pour emprunter des moniteurs de 21 po.

Les membres du personnel de production du secteur du recensement qui s'occupaient des aspects plus techniques ont offert d'évaluer les moniteurs et les cartes vidéo. On a réparti par roulement les moniteurs grands et moyens entre les intéressés pendant trois semaines. On a ainsi pu établir que, malgré une *capacité virtuelle* d'écran et une résolution de 1 280x1 024, les moniteurs de 17 po ne suffisaient pas à tous les besoins du recensement. Les cartes vidéo permettent des configurations virtuelles d'écran. Ainsi, l'utilisateur peut spécifier des aires d'affichage supérieures aux dimensions de l'aire d'affichage réelle de son moniteur. En maniant la souris, on peut faire apparaître les parties cachées de l'écran virtuel. On a jugé que cet écran était essentiel à certaines des activités sur moniteur de 21 po. Avec ces indications de recherche, on a pu arrêter des quantités de grands et de moyens écrans. Il a ensuite fallu passer aux considérations budgétaires.

Le prix, les dimensions et le rendement des moniteurs varient considérablement. On devait s'assurer que des moniteurs de 21 po pourraient être commodément disposés sur les bureaux, chose impossible dans certains cas. Aspect peut-être plus important encore, la qualité de l'affichage (netteté, contraste, luminosité et uniformité des images dans toute l'aire d'affichage) variait amplement pour des moniteurs d'une configuration et d'un prix comparables.

L'organisme maintient des offres à commandes (OAC) avec les fournisseurs en vue de faciliter ses acquisitions en technologie informatique. Sur la liste OAC, il y avait des appareils offerts par un certain nombre de fournisseurs. Aux fins du projet, on avait intérêt à exploiter cette liste si, parmi les moniteurs qui y figuraient, certains répondaient à nos besoins. Un représentant du groupe de soutien des postes de travail de l'organisme nous a fourni des appareils d'emprunt correspondant à tous les modèles de moniteurs 21 po de la liste. On a utilisé simultanément ces moniteurs à un même poste de travail et les mêmes images y ont été affichées. Le prix des appareils variait de 900 à 3 000 dollars canadiens. De prime abord, le moniteur le moins cher a bien paru, mais à y regarder de plus près, on a constaté que les couleurs se brouillaient et s'estompaient aux bords de l'écran. On a également observé un certain scintillement à une résolution de 1 024x1 280. Les moniteurs plus chers marchaient très bien, mais ils étaient d'un prix supérieur et d'une qualité inférieure par rapport à un des modèles déjà évalués par le personnel du recensement. On s'est arrangé avec le fournisseur pour qu'il ajoute ce produit à la liste OAC. (Les cartes vidéo désirées s'y trouvaient déjà.) Ces moniteurs ont été compris dans une commande de 120 Pentium destinés au personnel de production.

Pendant les activités de production, on a remarqué que les utilisateurs finals étaient nombreux à ne pas exploiter la capacité d'affichage virtuel des cartes vidéo simplement parce qu'ils ne s'y sentaient pas à l'aise. Ajoutons qu'une foule d'utilisateurs n'ont pas

utilisé les valeurs supérieures de résolution permises par les moniteurs et les cartes vidéo. Beaucoup ont choisi une résolution équivalente à celle de leurs anciens moniteurs SVGA de 14 po. Peut-être les petits caractères nets que l'on peut obtenir avec cette technologie sont-ils trop menus pour se lire facilement, surtout si les gens souffrent d'une déficience de la vue.

Contrôle de la qualité : Dans bien des systèmes d'imagerie, il y a des postes de travail et des préposés qui veillent sur la qualité des images communiquées au personnel de production. Les mesures de contrôle de la qualité (CQ) peuvent largement ajouter au coût de systèmes d'imagerie documentaire, car on doit souvent s'attendre à acquérir du matériel coûteux, à verser des salaires en plus et à subir des engorgements.

Les goulots d'étranglement sont particulièrement à craindre. La plupart des opérations en imagerie documentaire reposent sur une production d'images, et les opérations du recensement, sur une production de données. Ainsi, lorsqu'un vérificateur ou un analyste a accès aux données saisies dans la base de données structurées, il est prêt à commencer son travail et, si l'image du questionnaire correspondant n'a pas encore été communiquée, il travaillera sans cette image (auquel cas le système ne sert plus à rien) ou attendra qu'elle soit produite. Aux fins du recensement, il est impérieux d'obtenir les images ***avant les données structurées ou en même temps***. Pour pouvoir examiner (en temps utile) les quantités quotidiennes d'images produites par les lecteurs sans retard de production, il aurait fallu disposer d'un système coûteux de contrôle de qualité.

Par des discussions avec la direction du recensement, les agents principaux et les spécialistes des méthodes et des systèmes, on a cherché des façons de contourner le problème. Tant qu'une page restait lisible, peu importait à l'utilisateur si le reste du document ne l'était pas. Le plus souvent, l'utilisateur final ne s'intéressait qu'à une partie de page de chaque document de 16 pages qu'il demandait. Un spécialiste des méthodes a examiné les données d'évaluation des lecteurs automatiques (statistiques sur les cas de captation en double et les cas d'obliquité, de déchirure et de pliure des images) et prévu le nombre de fois qu'un utilisateur final rencontrerait une image inacceptable et demanderait une nouvelle mise en lecture. Ce nombre était négligeable, car les images produites par les lecteurs étaient invariablement d'une grande qualité. C'est ainsi qu'on a décidé de contrôler la qualité après les activités de production, le temps étant un facteur critique en production. La décision était d'autant plus facile à prendre que, pour demander une nouvelle lecture, il suffisait à l'utilisateur de ***cliquer sur un bouton***.

Comme on a fait preuve d'imagination et s'est attaché aux besoins plutôt qu'aux solutions techniques classiques, on a largement pu réduire l'ampleur et le coût du contrôle de la qualité. On n'a pas eu besoin de postes de travail supplémentaires, et les préposés ont fait du contrôle de qualité pendant les ***heures de relâche en production***.

Indexation : Comme dans d'autres types de bases de données, on emploie une ou plusieurs fonctions en relation avec les images pour une extraction en temps utile par l'utilisateur final. Dans certaines applications, on extrait automatiquement par lecture optique (reconnaissance optique de caractères ou ROC) les données d'identification des images avec lesquelles on dresse ensuite des index.

Comme les données d'identification des questionnaires étaient librement consignées par les recensés (elles n'étaient donc pas dactylographiées ni produites par imprimante), la ROC n'avait rien d'une solution applicable au recensement canadien de l'agriculture. On avait déjà écarté la RIC. Une autre possibilité était de se doter d'une application d'indexation interactive où le personnel devrait visualiser certaines pages de chaque image de questionnaire et saisir les données d'identification ensuite indexées (documents visualisés). Il y avait à cela un double inconvénient, puisque (1), avec des postes d'indexation, on se trouvait à accroître le coût du système d'imagerie et (2) on risquait de créer des goulots d'étranglement avant même les activités de production.

On pouvait heureusement obtenir les index nécessaires ailleurs et sans alourdir le budget. On avait tous les codes appropriés de renvoi à la base de données saisies des questionnaires. Il fallait une liaison entre le système d'imagerie et cette base d'information. On a demandé aux représentants de la Division du développement des systèmes de Statistique Canada d'étudier une solution.

On savait que les lecteurs pouvaient lire les codes à barres et les enregistrer automatiquement comme éléments d'indexation des images. Suivant les plans de saisie des données, des étiquettes de code à barres devaient être apposées aux questionnaires lors de l'introduction des données et saisies avec les données

des recensés. Ce serait là le code unique d'indexation versé dans la base de données.

En fonction de ce scénario, les spécialistes des systèmes ont proposé d'élaborer un programme permettant a) de consulter les nouvelles images de la base d'imagerie pour en extraire un code à barres et un code propre à cette base et b) de repérer, par le code à barres, l'enregistrement correspondant de la base de données structurées et y stocker le code interne de la base d'imagerie. L'utilisateur final aurait ainsi la possibilité d'accéder à un enregistrement de la base de données structurées par une des fonctions prédéfinies, puis de sélectionner l'image correspondante de la base d'imagerie à l'aide du code interne de cette dernière. C'était là une solution efficace et relativement peu compliquée.

Forme du système : Ce qui était très obscur dans la solution de l'imagerie documentaire était en train de se dissiper. Comme on n'avait besoin ni d'applications CQ, ni d'applications d'indexation, ni d'opérations de numérotation des questionnaires par gaufrage, on semblait se retrouver avec les seuls logiciels essentiels suivants :

- logiciel de gestion de lecture;
- logiciel de gestion de la base d'imagerie;
- programme de visualisation;
- logiciel de raccordement du système d'imagerie et de l'application du recensement (intégration).

Le matériel nécessaire était le suivant :

- lecteur(s);
- serveur;
- unité(s) de stockage;
- infrastructure réseau;
- postes de travail, moniteurs et cartes vidéo.

Quant à l'élément papier (questionnaire), on avait réglé la question de la couleur et de l'*adaptation au système d'imagerie*.

Logiciel de gestion de lecture automatique : On avait beaucoup appris au sujet de la fonctionnalité et du rendement des lecteurs par le projet pilote, les salons spécialisés et les discussions avec l'industrie, mais les évaluations avec le papier du recensement n'avaient pas visé un des chefs de file de l'industrie, la société KODAK. L'oubli semblait de taille, car KODAK avait autant de chances d'être choisie comme soumissionnaire DDP que les autres fabricants de lecteurs automatiques. On n'avait pas arrêté les détails essentiels (on savait seulement que le code à barres serait le code d'identification) du mode de gestion des documents du recensement mis en lecture.

Comme les modèles KODAK étaient relativement hauts de gamme par rapport aux lecteurs du projet pilote, on aurait sans doute besoin de moins de préposés à la lecture automatique. Un représentant de KODAK et un gestionnaire du recensement chargé de l'opérationnalisation des fonctions de lecture ont eu un entretien. Ils ont organisé l'emprunt de l'ImageLink 500 qui serait utilisé dans les locaux du recensement. Les représentants du secteur du recensement, un spécialiste des méthodes, des spécialistes des systèmes, des membres du groupe de soutien en micro-informatique et les représentants de KODAK ont travaillé en étroite collaboration pendant un mois. Il s'agissait non seulement d'apprendre à connaître la fonctionnalité et le rendement des modèles KODAK, mais aussi de comparer ces appareils à ceux des autres fabricants qui avaient été employés dans le projet pilote. On sentait de plus en plus quelle technologie de lecture automatique répondrait le mieux aux besoins du recensement.

On a conçu les essais pour qu'ils portent sur tous les aspects imaginables de la lecture automatique des documents du recensement. La gestionnaire des opérations, son adjointe et un spécialiste des méthodes :

- se sont attachés au débit des lecteurs;
- ont étudié la qualité des images;
- ont jugé si une lumière blanche suffisait au balayage des deux couleurs du questionnaire;
- ont appris ce qu'il fallait faire pour assurer le fonctionnement et l'entretien des lecteurs.

Voici les données primordiales qui ont été tirées de cet exercice :

(1) spécifications fonctionnelles des lecteurs à inclure dans la DDP;
(2) chiffres d'analyse comparative (nombre de cas de captation en double, d'obliquité des images et de blocage);
(3) tailles moyennes d'images KODAK acceptables (tailles qui s'accordaient avec celles des images produites dans le projet pilote);
(4) précisions sur le plan des besoins en dotation;
(5) constatation que le mécanisme de transmission KODAK travaillait mieux avec des demi-lots (15 documents de recensement) qu'avec des lots entiers (30);
(6) constatation que les codes à barres doivent être très lisibles pour que le lecteur les reconnaisse. Ces codes ne sont pas tous d'une qualité égale. Sans cette évaluation pratique, on aurait peut-être

fait cette constatation essentielle à un moment où tout aurait été paralysé.

À la suite de cet exercice, les responsables de la lecture automatique se sont attaqués aux difficultés considérables de l'opérationnalisation des fonctions de lecture. Il fallait notamment concevoir des opérations de réception, de tri, de balayage et de stockage éventuel des documents. On a ainsi conçu et spécifié un système automatisé de contrôle du cheminement des documents en temps normal et en cas de demande de nouvelle lecture. Un spécialiste des systèmes de l'organisme a construit ce système en Visual Basic, a participé aux essais dont il a fait l'objet et l'a installé. En production, celui-ci a été très efficient et d'un fonctionnement très sûr.

Pendant l'évaluation KODAK, un gestionnaire principal des tâches de recensement s'est occupé de l'élaboration et de la diffusion de la DDP. Pour ce faire, il a eu recours à la gestionnaire des opérations, à son adjointe, à l'agent de liaison technique du recensement, à un membre du groupe de soutien aux ordinateurs de gamme moyenne et à un représentant de la Division du développement des systèmes. La demande de propositions devait comprendre les éléments suivants :

1. renseignements généraux sur le recensement;
2. renseignements sur l'infrastructure technique (technologie serveur, système d'exploitation, protocole réseau, configuration et système d'exploitation d'informatique de bureau, etc.);
3. critères de fonctionnalité et de rendement des lecteurs;
4. besoins de formation;
5. besoins d'intégration;
6. contraintes techniques relatives à la taille des images;
7. exigences fonctionnelles de visualisation;
8. besoins de soutien;
9. critères d'acceptation;
10. délais.

Pour lancer une DDP dans la fonction publique fédérale, il faut suivre un grand nombre de règles et de règlements. Quelque nécessaire qu'il soit, cet exercice demande d'ordinaire jusqu'à six mois entre la remise d'une DDP aux Travaux publics et l'officialisation du choix d'un soumissionnaire. Le temps devenait un facteur critique dans le projet, les activités de production devant débuter dans moins de huit mois. Nous croyons que le processus DDP aurait pu compromettre la réalisation du projet si le responsable n'était pas toujours resté en contact avec les Travaux publics et n'avait pas fait pression sur ce ministère pour que la DDP du recensement demeure hautement prioritaire.

Visualisateurs : On se retrouvait devant l'alternative de faire l'acquisition de visualisateurs ou d'en construire sur mesure. Certains représentants de l'industrie pensaient qu'on économiserait considérablement en prenant l'option «visualisateurs sur mesure». D'autres étaient contre à cause des problèmes que posent les systèmes maison. On ignorait si le soumissionnaire retenu créerait ses propres visualisateurs ou en utiliserait qui existent déjà. D'une manière ou d'une autre, il fallait établir les spécifications fonctionnelles de visualisation et les inclure dans la DDP.

Cette détermination s'est révélée relativement simple. On a obtenu un visualisateur Watermark et l'a mis à la disposition de certains préposés au recensement. Après une démonstration et une période d'évaluation du logiciel, on a procédé à une «revue générale», occasion pour les intéressés d'indiquer simplement ce dont on avait besoin à leur avis et ce qu'il fallait modifier pour ensuite dégager des priorités. Les agents ont également précisé ce qu'ils jugeaient inutile. Les résultats de la consultation ont été documentés et intégrés à la DDP. On s'assurait ainsi que les besoins en visualisateurs seraient comblés par des produits standard ou une technologie sur mesure conçue par le sous-traitant.

Besoins de stockage : Cette détermination a été relativement simple. Après avoir examiné des échantillons d'images obtenues dans le projet pilote et dans l'évaluation des modèles KODAK, on a établi une taille moyenne d'image d'environ 600 K, valeur qu'on a multipliée par le nombre de questionnaires (en prévoyant une marge de sûreté de 25 %) pour arrêter l'ordre de grandeur du besoin global de stockage.

Il y avait toutefois là un certain risque, car on devait lancer la DDP de la technologie de stockage avant d'avoir choisi un soumissionnaire pour la DDP du système d'imagerie. La taille de l'image influait directement sur trois aspects primordiaux :

1. ***Qualité des images :*** Plus une image est grande, plus elle est de qualité. Il fallait veiller à ce que les images soient acceptables.
2. ***Besoins de stockage*** : Avec de grandes bases d'imagerie (4,8 millions d'images), une infime variation de la taille des images peut avoir une nette incidence sur les besoins de stockage. Il

fallait disposer d'une capacité de stockage suffisante avant les activités de production.

3. ***Rendement*** : Plus une image est grande, plus la charge est lourde pour le réseau, le serveur et le poste de travail. On avait fait tous les essais préalables de rendement en fonction d'une taille prévue des images. Si cette taille devait nettement augmenter, on aurait peut-être à en subir les conséquences.

Rendement : Un souci premier des gestionnaires de la production au recensement était celui de la vitesse de communication des images demandées aux ordinateurs de bureau. C'était peut-être là toute la différence entre la réussite et l'échec du système d'imagerie.

On avait pour défi de réduire les temps de communication au minimum. Le problème se posait lorsqu'il s'agissait d'une technologie à couches multiples qui transmettait les images et que d'autres activités se disputaient les capacités de calcul et d'acheminement réseau.

Voici les éléments de rendement que l'on devait examiner :

- Ordinateurs de bureau :
  - Vitesse de l'horloge;
  - Mémoire vive;
  - Antémémoire;
  - Espace de permutation;
  - Système d'exploitation (16 ou 32 bits);
- Moniteurs et cartes vidéo;
- Réseau :
  - Capacité (10 ou 100 Mo/s);
  - Protocole (TCP/IP, tailles des paquets, etc.);
  - Topologie (trafic isolé ou non);
- Images :
  - Taille des images;
  - Taille de la base d'imagerie;
  - Technologie de base de données;
- Serveur :
  - Mémoire vive;
  - Système d'exploitation;
  - Serveur réservé ou soutenant d'autres applications;
- Unités de stockage :
  - Antémémoire RAID;
  - Chargeurs automatiques de disques :
    - Densité optique des supports;
- Application d'imagerie :
  - Visualisateurs;
  - Gestion de base d'imagerie;
  - Index, etc.

Comme on pouvait s'y attendre, il était impossible de faire respecter des niveaux précis de rendement en production par le sous-traitant, et ce, à cause de toutes les conditions ambiantes qui restaient indépendantes de sa volonté. Il demeurait toutefois essentiel de veiller à ce que le rendement attendu du système du sous-traitant ne nuise pas à la réalisation du projet.

On a conçu des critères de rendement que l'on a intégrés à l'énoncé des critères d'acceptation de la DDP et mis en application. Dans un sous-réseau isolé, le sous-traitant a dû exécuter les tâches suivantes en employant la technologie de stockage et le serveur en production :

- mettre en lecture les versions du questionnaire du recensement de 1996;
- créer les images en reproduction multiple;
- garnir la base d'imagerie jusqu'à la moitié environ de la taille qu'elle aurait en production;
- à l'aide de deux postes de travail, extraire les quantités d'images représentatives prévues en production dans des délais précis; l'opération a été échelonnée (1, puis 10, puis 100 questionnaires, et ainsi de suite).

Dans ce scénario, le système du sous-traitant devait produire la première page d'un questionnaire en moins de 5 secondes et les pages suivantes en moins de 1 seconde.

Tout aussi importante était l'intégration de ce système aux systèmes de production créés à l'interne. Pour répondre aux critères, le sous-traitant a dû travailler avec un spécialiste de l'élaboration de systèmes et démontrer clairement que, avec l'application du recensement, on pouvait bien faire fonctionner un visualisateur (et produire des images).

À l'issue de ces essais, les représentants du secteur du recensement étaient raisonnablement sûrs que, si des problèmes de rendement devaient se poser, ce ne serait pas à cause du système du fournisseur.

Du côté «recensement» de la technologie, on a pris des mesures d'optimisation du rendement.

1. On a configuré, dimensionné et bien vérifié le réseau;
2. On a fait l'acquisition d'une grande antémémoire RAID pour y stocker les images fréquemment demandées;
3. On a acquis des Pentium avec assez de mémoire vive et d'espace disque;

4. On a aussi fait l'acquisition de cartes vidéo et de moniteurs à grand rendement.

Le résultat net a été une transmission satisfaisante d'images en production. Comme les temps d'arrêt ont été infimes, on a également constaté un haut degré de disponibilité du système.

Facteurs de succès : À l'époque où on a entrepris de se doter d'un système d'imagerie documentaire aux fins du recensement de l'agriculture, les exemples de catastrophes abondaient pour de tels projets, Statistique Canada ne comptait pas d'applications importantes d'imagerie et les gens chargés de mettre en place un tel système ne connaissaient absolument rien à cette technologie. L'exercice aurait pu se solder par un échec retentissant et fort coûteux pour le recensement canadien de l'agriculture.

On peut dégager un certain nombre de facteurs de réussite :

- ***Organisation de Statistique Canada en informatique.*** On avait sous la main pour le projet des compétences techniques en ordinateurs de gamme moyenne, en réseaux, en postes de travail et en élaboration de systèmes. Cette contribution a été aussi primordiale à l'étape de la production que dans les activités préparatoires.
- ***Infrastructure technique de Statistique Canada :*** Depuis quelque temps, l'organisme insistait fortement sur le respect des normes, ce qui a permis d'implanter dans cet environnement technique une technologie respectueuse des normes établies.
- ***Spécialiste des méthodes :*** Un spécialiste des méthodes a aidé les membres de l'équipe à analyser la technologie et à produire des statistiques utiles.
- ***Liaison technique :*** Aujourd'hui, les projets d'information font souvent appel à plusieurs disciplines techniques (ordinateurs de gamme moyenne, réseaux, postes de travail, élaboration d'applications, bases de données, produits, etc.). Dans ce projet d'imagerie, l'agent de liaison technique a coordonné les activités et représenté le secteur du recensement auprès de l'industrie et des groupes techniques.
- ***Système clés en main :*** Plutôt que de diviser un contrat entre plusieurs entreprises respectivement chargées de produire les images, de gérer le système d'imagerie et de fournir des services de visualisation, on a fait l'acquisition en bloc d'un système d'imagerie, d'où des garanties d'interfonctionnalité de tous les éléments du système (ainsi, les images produites par les lecteurs seraient compatibles avec les visualisateurs). On prévenait en outre les problèmes de communication dans les interventions d'entretien. Pour le secteur du recensement, il n'y avait qu'un sous-traitant à qui s'adresser.
- ***Simplicité :*** L'absence de postes d'indexation et de contrôle de qualité a rendu l'application moins complexe. Ajoutons que l'existence d'une interface minime entre l'application de production du recensement et le système d'imagerie FileNet a réduit les éléments et donc les problèmes d'intégration.
- ***Expérience directe :*** Il aurait été dangereux de s'en remettre à la documentation spécialisée et aux démonstrations, plus particulièrement en ce qui concerne la fonctionnalité des lecteurs automatiques. De plus, une évaluation pratique des moniteurs a permis de s'assurer d'une qualité suffisante des appareils. Il aurait été facile de trop dépenser ou encore de ne pas dépenser assez, ce qui aurait été une catastrophe. Dans la mesure du possible, le personnel du recensement a emprunté des modèles et en a fait l'expérience directe avec les documents et l'environnement du recensement.
- ***Primauté des besoins :*** Souvent dans ce projet, il aurait été plus facile de s'en remettre aux solutions classiques (postes d'indexation et de contrôle de qualité, etc.). De la sorte, on aurait inutilement augmenté les coûts, la complexité et les risques.
- ***Persévérance :*** Le projet d'imagerie aurait pu tourner court à plusieurs reprises.
- ***Soutien et participation directe de la direction :*** On sait fort bien qu'une foule de projets échouent par manque d'engagement de la direction. Le projet d'imagerie n'a pas souffert des conséquences d'un tel manque de participation des gestionnaires, loin de là. Le directeur adjoint et plus particulièrement le chargé de projet ont manifesté ouvertement leur appui, livré les batailles politiques et apporté une contribution de taille à la création du système.
- ***Relations avec le sous-traitant*** : Les rapports ont été excellents. La disponibilité, l'accessibilité et la disposition à faire des compromis ont été autant de facteurs de réussite. Il convient de noter que, avec ce projet, le sous-traitant s'est ouvert des portes dans l'administration fédérale. Le succès du projet était aussi important pour lui que pour le secteur du recensement, ce qui a sans doute joué dans les relations des deux parties.

- ***Abondance des contributions :*** Bien des gens ont été associés à la création du système d'imagerie. Les participants ont eu pour tâches :
  - d'évaluer les moniteurs, les cartes vidéo, les visualisateurs et les lecteurs;
  - de dimensionner le réseau et d'élaborer des programmes d'essai;
  - de définir la structure d'ensemble;
  - d'élaborer des DDP pour les essais en production et en projet pilote;
  - de définir comment les images seraient exploitées en production.
  - D'autres ont livré les inévitables batailles politiques.

  Ce haut degré de participation a suscité un plus grand sentiment d'«appartenance» et explique sans doute en partie la réussite du projet.
- ***Budget :*** Les frais d'acquisition, d'implantation et d'entretien ne sont pas négligeables. Une solide caisse noire nous a permis de donner les coups de barre qui s'imposaient en période d'implantation.
- ***Les questionnaires sur papier n'étaient plus une possibilité à envisager :*** Souvent, lorsqu'on introduit des technologies, les systèmes techniques à remplacer sont toujours là. Même si on a d'impérieuses raisons de vouloir changer les choses, les gens continuent dans bien des cas à faire leur travail avec les outils qui leur sont familiers jusqu'à ce qu'ils aient l'occasion de voir la nouvelle technologie. Pour le recensement de l'agriculture, l'application de 1996 représentait un changement radical. Le système manuel de remise de questionnaires qui avait été employé en 1991 n'existait plus. En 1996, il a fallu des circonstances exceptionnelles pour que l'on se résigne à produire un document sur papier.
- ***Acceptation :*** Le personnel de production s'est montré largement satisfait de la qualité des images et de la fonctionnalité des visualisateurs. Il a jugé extrêmement productives les fonctions d'annotation, de défilement et de zoom.
- ***Chance :*** Le projet a profité de nombreux talents dans l'équipe et les services. Deux ou trois des grands artisans du projet étaient remarquables.

Section III : Réflexions et perspectives d'avenir

On est parvenu au but fixé de transmission en temps utile d'images aux ordinateurs de bureau. Tous les utilisateurs et les observateurs ont bien accueilli - c'est le moins que l'on puisse dire - la technologie. L'imagerie et un excellent jeu d'instruments de vérification et de validation se sont révélés une combinaison gagnante pour les utilisateurs. Tout ce que se demandait le personnel, c'est comment on avait pu faire par le passé pour traiter les données du recensement sans imagerie.

On a obtenu les avantages prévus et plus encore. L'imagerie s'est implantée dans une foule d'activités et d'applications où elle n'était pas du tout attendue. Voici quelques retombées imprévues :

- Le gain de productivité a largement dépassé les 10 % prévus. Dans un grand nombre d'opérations par commis, le visualisateur a automatiquement fait un zoom sur la partie problème du questionnaire et la correction s'est faite en direct. On travaillait bien plus vite qu'en feuilletant des questionnaires et des imprimés d'ordinateur. Dans le système manuel, il fallait notamment écrire la correction sur l'imprimé pour que quelqu'un d'autre l'apporte ensuite en direct.
- En période de collecte, il est possible de recevoir plusieurs questionnaires de recensement d'une exploitation agricole. Il y a une marche à suivre pour constater cette réception multiple et prendre des mesures correctives. En consultant des images des questionnaires d'une même exploitation sur écran partagé, on épargne bien du temps.

- Dans cette même période de collecte, les problèmes de sous-dénombrement obligeaient les préposés à identifier les exploitations oubliées ou manquantes. Pour ce faire, il fallait comparer les listes du recensement de 1996, des services fiscaux, des enquêtes et des recensements antérieurs. Grâce aux images, les vérificateurs ont pu rapidement voir si une exploitation figurait dans la base de données. On a réduit et les coûts de suivi sur le terrain et le fardeau de réponse et les risques de dédoublement des questionnaires. Les images ont aussi donné des indications (descriptions des terres, etc.) qui ne comptaient pas parmi les données saisies dans la base de données structurées.

- Un service en croissance rapide dans le secteur canadien de l'agriculture est celui des élevages porcins à contrat. Nombre d'éleveurs à contrat n'ont pas déclaré leurs porcins sur leur questionnaire malgré les instructions données. Comme les porcs ne leur appartenaient pas, ils n'ont pas jugé bon de les inclure. Nous avions

donc des données sur les exploitations, mais sans les porcs. En comparant le recensement de 1996 à l'enquête semestrielle sur les porcins, nous avons pu voir quels exploitants indiquaient un grand nombre de porcs dans l'enquête, mais non pas dans le recensement. Les animaux manquaient-ils par erreur de traitement ou de réponse? On a découvert la solution en consultant rapidement les images. On a ensuite fait le suivi nécessaire auprès des recensés pour qu'ils ajoutent les porcs manquants.

- Avec les méthodes d'imputation automatisées, il semble toujours se présenter des situations où, pour chaque module de questions, il est impossible de trouver un enregistrement donneur. Fréquemment, des erreurs de réponse ou de traitement sont en cause. Grâce aux images, les agents spécialisés ont pu prendre des mesures correctives. Dans les recensements du passé, on apportait les corrections au jugé pour faire passer les questionnaires à la prochaine étape d'imputation. Si on avait pris le temps de repérer et d'extraire les documents sur papier, on aurait sérieusement retardé les opérations d'imputation.

L'imagerie a-t-elle fait ses frais en une seule période de recensement?
Non, et il ne fallait pas s'y attendre. Comment mettre une valeur économique sur la qualité des données, la satisfaction au travail ou le coût du report d'une date de diffusion?

Pourra-t-on amortir les frais du système d'imagerie à court terme (en cinq ans)? Oui, sûrement. Le matériel continue à être utilisé au recensement dans des essais, ainsi que dans bien d'autres projets en expérimentation ou en production. Il jouera aussi un rôle - avec des perfectionnements et des compléments - dans le recensement de 2001. L'expérience fructueuse qui en a été faite au recensement de l'agriculture a encouragé d'autres secteurs à essayer et à implanter non seulement la technologie d'imagerie, mais aussi d'autres applications RIC. Voici des secteurs qui s'occupent aujourd'hui de cette technologie :

- Projet des données complètes sur les exploitations agricoles;
- Recensement de la population;
- Division des données fiscales;
- Division de l'organisation et des finances de l'industrie;
- Division des opérations et de l'intégration;
- Division des services de soutien administratif (Gestion des documents);
- Projet d'amélioration des statistiques économiques provinciales (PASEP).

Parfois, le recensement de l'agriculture s'est chargé de travaux se rattachant à d'autres projets. Ce travail supplémentaire :

- aide à payer le système et permet de partager les frais d'entretien et de licence;
- soutient des recherches liées;
- entretient et enrichit l'expérience du personnel;
- ajoute des caractéristiques et des capacités au système.

Nul doute que la décision d'implanter un système d'imagerie a été la bonne. Le risque que l'on a pris était acceptable, mais non négligeable. S'il y avait eu de nouvelles difficultés, tout le projet aurait été en danger.

L'expérience prouve qu'il valait la peine de prendre un tel risque. Somme toute, cet élément de risque a joué un rôle primordial en aidant le secteur du recensement à respecter ses dates prévues de diffusion.

Divers événements avaient retardé les activités. À eux seuls, les problèmes de sous-dénombrement avaient fait accuser un retard de plusieurs semaines aux tâches de validation des données. De plus, les problèmes du système généralisé d'imputation avaient imposé une grande opération de colmatage et une réimputation de toutes les variables financières pour deux des principales provinces du pays.

Par d'exceptionnels efforts concertés des membres de l'équipe et grâce aux gains d'efficience permis par le système d'imagerie, les données du recensement ont pu être diffusées à temps.

Où la technologie mène-t-elle le recensement de l'agriculture de 2001? La technologie RIC peut-elle traiter des questionnaires de 16 pages sur papier ministre remplis par 300 000 exploitants agricoles qui gagnent en âge?

Là encore, on observe une diversité d'opinions et de camps. Il y a des membres de l'équipe de projet qui étaient conservateurs en 1996 et qui sont progressistes aujourd'hui et vice versa. D'autres choses ont changé aussi. Nous savons que l'imagerie fonctionne et donne de bons résultats. Nous avons plus d'expérience et de confiance, et plus de temps aussi.

La recherche RIC pour le recensement de l'agriculture a débuté il y a presque un an, soit quatre ans avant le recensement de 2001. Les premiers résultats sont certes encourageants, mais il est bien trop tôt pour établir des prévisions et encore plus pour prendre des décisions. C'est en 1998 qu'on commencera à évaluer la réaction des recensés aux questionnaires adaptés à la technologie RIC.

Quelle sera notre position de repli si la RIC est jugée peu rentable ou peu appropriée au recensement de l'agriculture? Une possibilité «conservatrice» serait l'imagerie avec plus de traitement en parallèle. Un autre grand pas en avant serait la «lecture automatique de marques») avec extraction-images.

Certains voient dans l'imagerie documentaire une technologie transitoire dans la voie menant au bureau sans papier. On peut se demander quand la déclaration électronique remplacera la déclaration sur papier aux fins du recensement de l'agriculture.

BIBLIOGRAPHIE

1. Bradshaw, C., et Duggan, J. (1997) Évaluation de l'imagerie documentaire appliquée au recensement canadien de l'agriculture de 1996.
2. Duggan, J. (1996) Electronic Imaging in Support of the 1996 Census of Agriculture, document présenté à la conférence internationale sur la collecte assistée par ordinateur des données des enquêtes.
3. Green, I. (1995) Questionnaire Imaging Pilot, rapport interne, Statistique Canada.

Série de documents de travail sur l'agriculture et le milieu rural

La Division de l'agriculture publie des documents de travail sur la recherche, les résultats analytiques, les techniques statistiques, les méthodes et les concepts. Pour commander, appelez la Division de l'agriculture sans frais en composant le 1-800-465-1991. (Certains des documents de travail sont disponibles sur Internet.)

N° 1	(21-601-MPF80001)	**Description de la méthode Theil de prévision de l'erreur quadratique moyenne pour la statistique agricole (1980)**	Stuart Pursey
N° 3	(21-601-MPF81003)	**Examen du Projet de l'estimation du bétail et recommandations de mesures à prendre (1981)**	Bernard Rosien et Elizabeth Leckie
N° 4	(21-601-MPF84004)	**Le secteur canadien des oléagineux : vue d'ensemble (1984)**	Glenn Lennox
N° 5	(21-601-MPF84005)	**Analyse préliminaire de la contribution des paiements directs du gouvernement dans le revenu agricole net réalisé (1984)**	Lambert Gauthier
N° 6	(21-601-MPF84006)	**Les caractéristiques des exploitants entrant en agriculture et leurs entreprises au sud de l'Ontario pour la période 1966 à 1976 (1984)**	Jean B. Down
N° 7	(21-601-MPF84007)	**Sommaire des programmes d'aide à la production agricole aux États-Unis (1984)**	Allister Hickson
N° 8	(21-601-MPF84008)	**Intensité de la pratique de la jachère dans les Prairies : Une analyse des données du recensement de 1981 (1984)**	Les Macartney
N° 9	(21-601-MPF85009)	**Évolution de la structure du secteur porcin au Canada (1985)**	Mike Shumsky
N° 10	(21-601-MPF86010)	**Révisions au traitement des loyers de maisons imputés dans les comptes de fermes canadiennes, 1926-1979 (1986)**	Mike Trant
N° 11	(21-601-MPF92011)	**L'estimateur par le quotient : explication intuitive et utilisation pour estimer les variables agricoles (1992)**	François Maranda et Stuart Pursey
N° 12	(21-601-MPF91012)	**L'effet de la distortion géographique causée par la règle de l'emplacement (1991)**	Rick Burroughs
N° 13	(21-601-MPF91013)	**La qualité des données agricoles : forces et faiblesses (1991)**	Stuart Pursey
N° 14	(21-601-MPF92014)	**Autres cadres d'examen des données rurales (1992)**	A.M. Fulier, Derek Cook et Dr John Fitzsimons
N° 15	(21-601-MPF93015)	**Tendances et caractéristiques relatives aux régions rurales et aux petites villes du Canada (1993)**	Brian Biggs, Ray Bollman et Michael McNames
N° 16	(21-601-MPF92016)	**La microdynamique et l'organisation économique de la famille agricole dans le changement structurel en agriculture (1992)**	Phil Ehrensaft et Ray Bollman
N° 17	(21-601-MPF94017)	**Consommation de céréales et de graines oléagineuses par le bétail et la volaille, Canada et provinces, 1992**	Section du bétail et des produits d'origine animale
N° 18	(21-601-MPF94018)	**Changements structurels dans le domaine agricole - Étude comparative des tendances et des modèles observés au Canada et aux États-Unis**	Ray Bollman, Leslie A. Whitener et Fu Lai Tung

N° 19	(21-601-MPF94019)	**Revenu total de la famille agricole selon le type d'exploitation et la taille de celle-ci, et selon la région, en 1990 (1994)**	Saiyed Rizvi, David Culver, Lina Di Piétro et Kim O'Connor
N° 20	(21-601-MPF94020)	**L'adaptation dans le secteur agricole au Canada (1994)**	George McLaughlin
N° 21	(21-601-MPF93021)	**Microdynamique de la croissance et de la décroissance des exploitations agricoles : une comparaison Canada - États-Unis**	Fred Gale et Stuart Pursey
N° 22	(21-601-MPF92022)	**Les structures des gains des ménages agricoles en Amérique du Nord – Positionnement pour la libéralisation des échanges**	Leonard Apedaile, Charles Barnard, Ray Bollman et Blaine Calkins
N° 23	(21-601-MPF92023)	**Secteur de la pomme de terre : comparaison entre le Canada et les États-Unis**	Glenn Zepp, Charles Plummer et Barbara McLaughlin
N° 24	(21-601-MPF94024)	**Étude comparative des données américaines et canadiennes sur la structure des fermes**	Victor J. Oliveira, Leslie A. Whitener et Ray Bollman
N° 25	(21-601-MPF94025)	**Méthodes statistiques de la Sous-section de la commercialisation des céréales, document de travail, version 2**	Karen Gray
N° 26	(21-601-MPF94026)	**Rendement des exploitations agricoles : Estimations établies à partir de la base de donnés complètes sur les exploitations agricoles**	W. Steven Danford
N° 27	(21-601-MPF94027)	**La mesure de l'emploi touristique dans les régions rurales**	Brian Biggs
N° 28	(21-601-MPF95028)	**Délimitation de l'écoumène agricole canadien de 1991**	Timothy J. Wershler
N° 29	(21-601-MPF95029)	**Étude cartographique de la diversité des économies rurales : une typologie préliminaire du Canada rural**	Liz Hawkins
N° 30	(21-601-MPF96030)	**Structure et tendances de l'emploi rural au Canada et dans les pays de l'OCDE**	Ron Cunningham et Ray D. Bollman
N° 31	(21-601-MPF96031)	**Une nouvelle approche pour les régions autres que les RMR/AR**	Linda Howatson-Leo et Louise Earl
N° 32	(21-601-MPF96032)	**L'emploi dans l'agriculture et ses industries connexes en région rurale : structure et changement 1981-1991**	Sylvain Cloutier
N° 33	(21-601-MPF98033)	**Exploiter une ferme d'agrément – pour le plaisir ou le profit?**	Stephen Boyd
N° 34	(21-601-MPF98034)	**Utilisation de la technologie d'imagerie documentaire dans le recensement canadien de l'agriculture de 1996**	Mel Jones et Ivan Green
N° 35	(21-601-MPF98035)	**Tendances de l'emploi au sein de la population active non métropolitaine**	Robert Mendelson
N° 36	(21-601-MPF98036)	**La population des milieux ruraux et des petites villes s'accroît pendant les années 90**	Robert Mendelson et Ray D. Bollman
N° 37	(21-601-MPF98037)	**La composition des établissements commerciaux dans les petites et les grandes collectivités du Canada**	Robert Mendelson
N° 38	(21-601-MPF98038)	**Le travail hors ferme des exploitants de fermes de recensement : Aperçu de la structure et profils de mobilité**	Michael Swidinsky, Wayne Howard et Alfons Weersink

N° 39	(21-601-MPE99039)	**Le capital humain et le développement rural : quels sont les liens?**	Ray D. Bollman
N° 40	(21-601-MPE99040)	**Utilisation de l'ordinateur et d'Internet par les membres des ménages ruraux**	Margaret Thompson-James
N° 41	(21-601-MPE99041)	**Les Cotisations aux réer des producteurs agricoles Canadiens en 1994**	Marco Morin
N° 42	(21-601-MPE99042)	**Intégration des données administratives et des données d'enquête de recensement**	Michael Trant et Patricia Whitridge

Prix:

1 copie	**Gratuite**
2 à 10 copies	**50,00 $**
11 à 25 copies	**75,00 $**
26 à 38 copies	**100,00$**